Journal of Medical Internet Research

Contents

Reviews

Viewpoints

Original Papers

XSL·FO
RenderX

XSL•FO
RenderX

Corrigenda and Addendas

XSL•FO
RenderX

The Effectiveness of Virtual Reality in Managing Acute Pain and Anxiety for Medical Inpatients: Systematic Review

Vinayak Smith[1,2*], MBBS, DPhil; Ritesh Rikain Warty[1,2*], BEng; Joel Arun Sursas[2], MBBS; Olivia Payne[1], BBSc; Amrish Nair[2], MEng; Sathya Krishnan[3], MBBS; Fabricio da Silva Costa[1,4], DPhil; Euan Morrison Wallace[1], MD; Beverley Vollenhoven[1], MBBS, DPhil

[1]Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia

[2]Biorithm Pte Ltd, Singapore, Singapore

[3]Department of Paediatrics, West Gippsland Hospital, Warragul, Australia

[4]Department of Gynecology and Obstetrics, Ribeirão Preto Medical School, Sao Paulo, Brazil

[*]these authors contributed equally

Corresponding Author:
Vinayak Smith, MBBS, DPhil
Department of Obstetrics and Gynaecology
Monash University
246 Clayton Road
Clayton
Australia
Phone: 61 431330754
Email: vinayak.smith@monash.edu

Abstract

Background: Virtual reality is increasingly being utilized by clinicians to facilitate analgesia and anxiolysis within an inpatient setting. There is however, a lack of a clinically relevant review to guide its use for this purpose.

Objective: To systematically review the current evidence for the efficacy of virtual reality as an analgesic in the management of acute pain and anxiolysis in an inpatient setting.

Methods: A comprehensive search was conducted up to and including January 2019 on PubMed, Ovid Medline, EMBASE, and Cochrane Database of Systematic reviews according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Search terms included virtual reality, vr, and pain. Primary articles with a focus on acute pain in the clinical setting were considered for the review. Primary outcome measures included degree of analgesia afforded by virtual reality therapy, degree of anxiolysis afforded by virtual reality therapy, effect of virtual reality on physiological parameters, side effects precipitated by virtual reality, virtual reality content type, and type of equipment utilized.

Results: Eighteen studies were deemed eligible for inclusion in this systematic review; 67% (12/18) of studies demonstrated significant reductions in pain with the utilization of virtual reality; 44% (8/18) of studies assessed the effects of virtual reality on procedural anxiety, with 50% (4/8) of these demonstrating significant reductions; 28% (5/18) of studies screened for side effects with incidence rates of 0.5% to 8%; 39% (7/18) of studies evaluated the effects of virtual reality on autonomic arousal as a biomarker of pain, with 29% (2/7) demonstrating significant changes; 100% (18/18) of studies utilized a head mounted display to deliver virtual reality therapy, with 50% being in active form (participants interacting with the environment) and 50% being in passive form (participants observing the content only).

Conclusions: Available evidence suggests that virtual reality therapy can be applied to facilitate analgesia for acute pain in a variety of inpatient settings. Its effects, however, are likely to vary by patient population and indication. This highlights the need for individualized pilot testing of virtual reality therapy's effects for each specific clinical use case rather than generalizing its use for the broad indication of facilitating analgesia. In addition, virtual reality therapy has the added potential of concurrently providing procedural anxiolysis, thereby improving patient experience and cooperation, while being associated with a low incidence of side effects (nausea, vomiting, eye strain, and dizziness). Furthermore, findings indicated a head mounted display should be utilized to deliver virtual reality therapy in a clinical setting with a slight preference for active over passive virtual reality for analgesia. There, however, appears to be insufficient evidence to substantiate the effect of virtual reality on autonomic arousal, and this should be considered at best to be for investigational uses, at present.

Introduction

Background

The International Association for the Study of Pain defines pain as an "unpleasant sensory and emotional experience associated with actual or potential tissue damage [1]." As a disease, pain carries a huge global burden with a prevalence of 20% and incidence of 10%. It negatively affects one's psychological and social functioning, thereby impinging on quality of life too. Its tangible costs cannot be understated as well, with research demonstrating a loss in productivity owing to absenteeism and diminished job performance as a result [2-7].

Acute pain is a commonly encountered clinical entity in up to 84% of patients presenting to medical services [8]. Acute pain is sudden in its onset and is typically expected to last for a short time (≤6 weeks). Usually, it can be attributed to a specific event or illness, but at times, it may be iatrogenic [9].

Either pharmacologic (analgesics) and nonpharmacologic (interventions) can be used for addressing acute pain. Within analgesics, opioids are often prescribed [10]. This strategy, although clinically effective, has several disadvantages. Opioids are notorious for their deleterious side effects, including tolerance, dependence, and hyperalgesia [11]. Additionally, some analgesics also require invasive procedures to be administered, such as with intrathecal infusions, which carry their own set of clinical risks and side effects. Similarly, concerns have been raised toward nonpharmacologic approaches (ie, transcutaneous electrical nerve stimulation, hot or cold compress), with regard to their efficacy and appropriateness in the setting of acute pain [10,11].

In light of these findings, as well as the recent recommendations from the American Pain Society and the American Society of Anaesthesiologists, there remains an urgent need to characterize and identify alternative modalities for acute pain relief. In particular, there is a need for these therapies to be clinical efficacious, be minimally invasive, and potentiate low levels or negligible side effects [12]

Virtual reality is a burgeoning technology which is in its infancy of uptake for clinical utilization. As a technology, virtual reality allows for users to be immersed in a virtual environment through multisensorial stimulation [13,14].

Over the last decade, increasing attention and research has been directed toward assessing the utility of virtual reality in managing acute pain. While there is presently no clear explanation of virtual reality's mechanism of pain relief available, several theories which span the realms of biology and psychology exist to elucidate its efficacy [15,16]. To date, it is also worthy of mention that virtual reality therapy has been successfully used as an analgesic in several acute clinical contexts, ranging from pediatric phlebotomy to dressing changes for burns and postcardiac surgery [17-19].

Objectives

The main motivation for our group in undertaking this systematic review was to provide a comprehensive literature review to inform the clinical utilization and testing of virtual reality therapy.

Primarily, this encompassed understanding the applicability of virtual reality in facilitating analgesia during acute pain for inpatient populations. This also included an understanding of virtual reality therapy headsets, content being used for the indication, and the effects of virtual reality on anxiolysis, since it has been implicated in facilitating this effect, which in turn, modulates the patient's perception of pain [20,21]. Furthermore, we aimed to delineate the effect of virtual reality on physiological parameters (autonomic arousal); literature has suggested that these are biomarkers of pain, and thus, are theoretically correlated with pain responses [22]. Last, given the impact that it would have on patient safety and clinical uptake, we wanted to understand virtual reality therapy's side effect profile.

Prior to undertaking this review, there was a gap in the literature on virtual reality therapy, in the setting of facilitating inpatient acute analgesia, which was of practical relevance to the clinician. This, in our opinion, appeared to be a barrier of clinical uptake which we aimed to address through this initiative by providing a holistic overview.

Methods

Data Sources

The following review was conducted in line with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The search was undertaken on PubMed, Ovid Medline, EMBASE, and the Cochrane Database of Systematic Reviews up to and including May 2018 by VS, JS, AN, and SK and repeated up to and including January 2019 by VS, RRW, OP, and JS. The search was carried out without any limit of the years, and articles were restricted to those in English. The databases were searched independently by the aforementioned authors. Once shortlisted, full texts were ordered and read. The bibliographies of articles selected for the review were also screened for suitable additional articles to be included in this review. Inclusion in the review was selected by consensus between the screening authors.

Inclusion criteria were primary studies utilizing virtual reality in the management of acute pain in a clinical setting. Interventions were considered to be virtual reality only if they employed an audio or visual multimedia environment with which the patients could view or interact (ie, games and videos). Acute pain was defined as pain that was less than 6 weeks in duration and associated with an acute condition or medical procedure. The context was selected to make the findings relevant to the inpatient treatment of patients for clinicians.

Studies with both adult and pediatric populations were suitable for inclusion in the review.

Articles that were reviews, case series, or case reports were excluded from this review. Experimental studies of a nonclinical nature (eg, pain induced via cold-pressor test) were excluded in a bid to focus on clinically relevant pain reduction which could be easily extrapolated to clinical practice.

Search Strategy

Search terms used across all databases were (*virtual reality* OR *vr*) AND *pain*. Studies were then filtered manually as per the inclusion criteria for acute pain associated with acute conditions or procedures.

Data Collection Process

Data were extracted manually for analysis by VS, SK, and JA in tabular form. Due to the heterogeneity of the studied populations, variations in technologies utilized, and heterogeneity in the endpoints of the studies; pooling of data for meta-analysis was not considered appropriate. In addition, meta-analysis of data from rapidly evolving medical

technologies of various generations was deemed inappropriate [23]. This is due to the lack of similarity between technologies, the impact of incremental innovation between generations of the same technology and the presence of operator dependence (from a clinician and patient perspective) on its performance [23]. As such, a narrative approach was followed for this review.

Data Items

Data items of interest for the studies included: year of study, study design, sample size, clinical setting, population, nature of the intervention, control or comparison, virtual reality content type, main outcomes measures, outcome measurement tools, and technical specifications of virtual reality devices employed. Technical specifications of virtual reality devices included type of head mounted display, display utilized, weight of device, field of view, computer, video card, controller, virtual reality content type, and virtual reality content used.

Assessment of Bias

The risk of bias was assessed by VS and RRW using the modified Downs and Black List [24] and scored on a scale of 1-10 as illustrated in Table 1.

Table 1. Summary of bias assessment results using modified Downs and Black checklist.

Study	Score (out of 10)
Chad et al [25]	9
Chan et al [26]	6
Chau et al [27]	8
Frey et al [28]	9
Gerceker et al [29]	10
Gershon et al [30]	8
Glennon et al [31]	9
Gold et al [32]	9
Hoffman et al [18]	8
McSherry et al [33]	7
Mosso-Vasquez et al [17]	8
Mosso-Vasquez et al [34]	7
Nilsson et al [19]	9
Piskorz et al [35]	8
Shoorab et al [36]	9
Tashjian et al [37]	9
Walker et al [38]	8
Yun Hua et al [39]	9

Summary Measures and Synthesis of Results

The primary measures of interest were degree of analgesia afforded by virtual reality therapy, degree of anxiolysis afforded by virtual reality therapy, effect of virtual reality on physiological parameters, side effects precipitated by virtual reality, measures of pain assessment, virtual reality content type, and types of equipment utilized.

Patient and Public Involvement

No patients were involved in the design, recruitment, or conduct of the study. There was no intention a priori that the results of this review would be disseminated to patients included in the trials of the review.

Results

Study Characteristics

General Description of Studies

A total of 18 studies were deemed suitable for inclusion in this review [17-19,25-39]. The article selection process is outlined in Figure 1.

The descriptive data collated from the eligible studies are reported in Table S1 (Multimedia Appendix 1). All 18 studies were conducted and published between 2004 and 2018: 50% (9/18) of studies were specifically focused on the pediatric population; 72% (13/18) of studies compared virtual reality against standard analgesia as the comparator/control group [18,19,26,29-36,38,39]; 28% (5/18) of studies compared virtual reality to no analgesia [17,25,27,28,37]. Overall, only 1 study of 18 (6%) received a bias assessment score less than 7 out of 10.

Figure 1. PRISMA flow diagram.
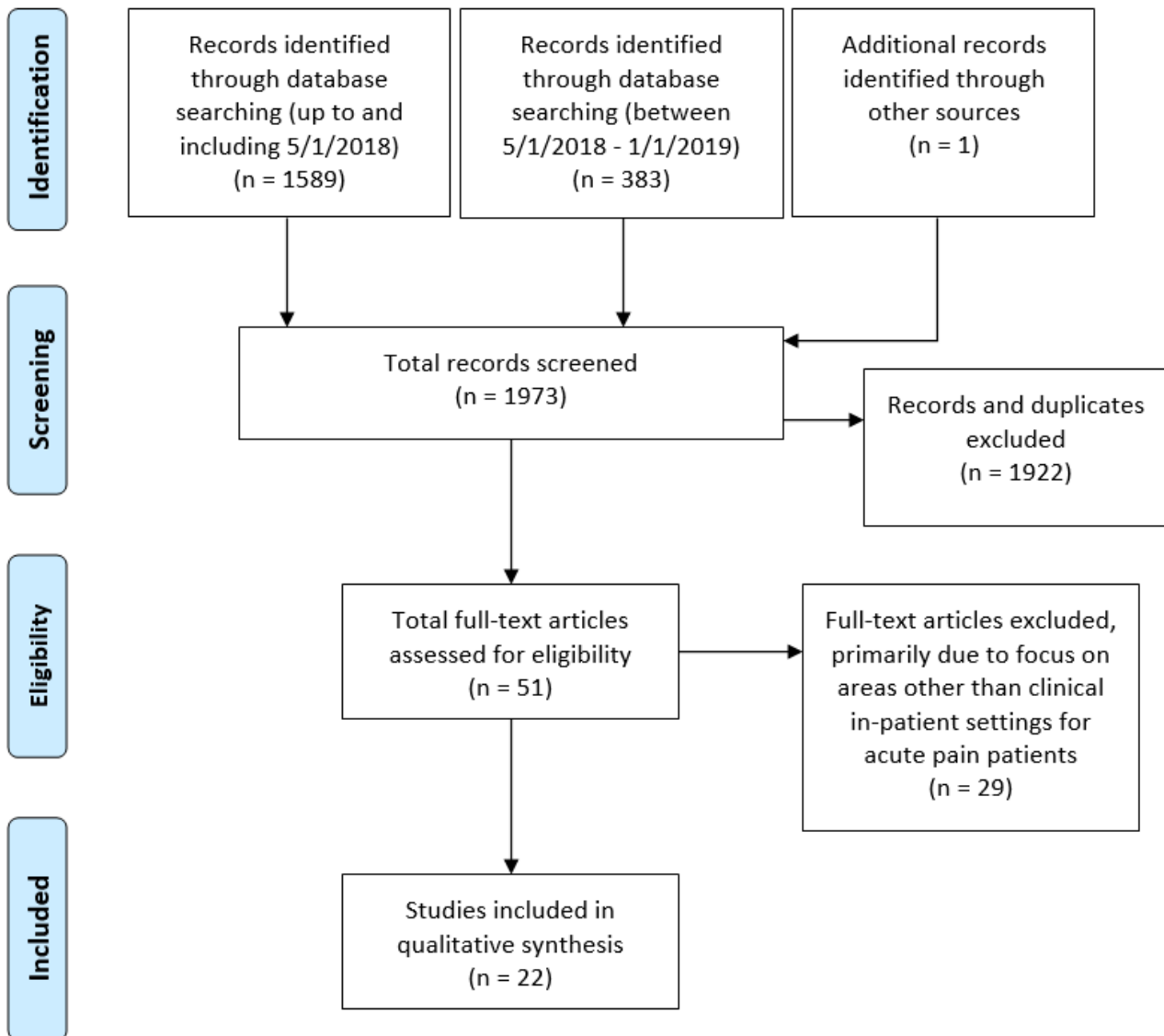
Indications for Virtual Reality Utilization

Of the 18, 22% (4 studies) measured pain reduction in patients undergoing dressing changes or hydrotherapy for burns or wounds [18,26,33,39]; 39% (7 studies) looked at patients undergoing venipuncture, port access, or injections [19,25,27,29,30,32,35]; 39% (7 studies) measured pain outcomes in patients undergoing various invasive procedures (5 studies) [17,31,34,36,38] or other miscellaneous acute conditions (2 studies) [28,37].

Technical Specifications of Virtual Reality Systems Utilized

Table S2 (Multimedia Appendix 2) summarizes data on the technical specifications of devices.

All studies (18/18, 100%) utilized head mounted displays which allowed for monitoring of head orientation. Half (9/18, 50%) used active virtual reality content in the form of games [18,19,26,30,32,33,35,38,39] as the primary content, and the remaining 50% of the studies (9/18) employed passive virtual reality content in the form of virtual environments and movies.

One-third (6/18, 33%) applied a virtual environment [17,25,28,31,34,37], and 17% (3/18) displayed a cartoon or movie [27,29,36].

Summary of Measures and Synthesis of Results

Effect on Analgesia

Overall, 67% (12/18) of studies in this review demonstrated a statistically significant reduction in pain during virtual reality utilization; 83% (10/12) of these demonstrated a statistically significant reduction in pain using a between-group design [18,26,28-30,32,35-37,39]. Within these 10 studies, only 1 received a score less than 7 on the bias assessment checklist. The remaining 17% (2/12) of studies demonstrated statistically significant reductions in perceived pain utilizing a within-group design [25,34]. Across studies, the clinical endpoint of pain was measured through a variety of pain measurement tools. These results and tools are described in Table 2.

Table 2. Summary of studies for utility of virtual reality in acute pain.

Study	Intervention	Standard care; control	Measurement tools	Results
Chad et al [25]	VR[a]	N/A[b]	FACES[c] (scale of 0-5)	Descriptive evidence of significantly reduced pain as reported by parents: 83% reduction (P=.02) in pain noted pre-VR (mean 3.34, SD 1.30) and post-VR (mean 0.76, SD 1.06).
				Insignificant reduction in pain reported by children: 77% reduction in pain noted with VR use (mean difference –2.5; P=.52).
				No information on pain score from children and variance.
Chan et al [26]	VR	Routine analgesia	FACES (scale of 0-100)	Significant differences (P<.05) of pain intensity found between VR group and non-VR group with ANOVA[d]: pain during procedure (VR: mean 38.13, SD 12.02; control: mean 53.75, SD 11.80) and pain after procedure (VR: mean 8.75, SD 2.95; control: mean 18.75, SD 9.53).
Chau et al [27]	VR	N/A	FLACC[e] (scale of 0-10)	FLACC scores with VR use: median 2.5 (1-5.5); range (1-8).
				64.3% (9/14) of caregivers felt that the VR experience was positive and wanted to use it again. No between group differences explored for the following study.
				Descriptive evidence of reduced pain as reported by parents.
Frey et al [28]	VR	Unmedicated labor	VNS[f] (scale of 0-10)	Worst pain intensity was significantly lower in VR vs control (–1.5 (95% CI, –0.8 to –2.2).
				Significant differences in difference of pain intensity found between VR group and non-VR group with ANOVA.
Gerceker et al [29]	VR	External cold and vibration (buzzy); analgesia alone	Wong Baker FACES (scale of 0-10)	There was a statistically significant difference observed in pain scores between the VR group and control as reported by the patients (VR: mean 1.5, SD 0.2; control: mean 5.1, SD 0.4; P<.01) and parents (VR: mean 1.5, SD 0.2; control: mean 4.7, SD 0.4; P<.01).
Gershon et al [30]	VR with topical anesthetic	Video game with topical anesthetic; topical anesthetic	VAS[g] (scale of 0-100); CHEO[h] pain scale	There was a significant reduction in pain in the VR and non-VR distraction group as observed by nurses (P<.05) based on MANOVA[i]. No summary measures were provided in the paper.
				From the CHEO pain scale, the VR group had significantly fewer behavioral markers in comparison to those of the controls for pain (P<.05). No summary measures were provided in the paper.
				No summary statistics on pain score from children and variance.
Glennon et al [31]	VR	Analgesia alone	NPS[j] (scale of 0-10)	Participants in the VR group (mean 3.9, SD 2.3)did not experience a statistically significant decrease (P<.05) in pain and anxiety in comparison to controls (mean 4, SD 2.7).
				Not powered adequately. Recruited individuals with prior exposure to bone marrow biopsy which may have skewed the reporting of pain.
Gold et al [32]	VR and analgesia	Analgesia alone	VAS (scale of 0-10); CAS[k]; FACES	Significantly less pain (P<.05) was reported by the VR group (mean 1.31, SD 1.59) compared to that reported by the control group (mean 1.93, SD 2.22).
Hoffman et al [18]	VR and analgesia	Analgesia alone	GRS[l]	Statistically significant reduction (P<.01) in pain observed in VR (mean 7.6, SD 1.9) vs control (mean 5.1, SD 2.6).

XSL•FO
RenderX

Study	Intervention	Standard care; control	Measurement tools	Results
McSherry et al [33]	VR and analgesia	Analgesia alone	VNS; opioid administration	Total opioid administration during VR therapy was significantly less than that when no VR was used (VR: 91.7 SD 10.1; no VR: 103.1 SD 16.1 µg/kg; $P<.05$). Overall, 39% less opioids were used with VR therapy.
				Pain scores (post procedure – pre procedure) were not significantly ($P>.05$) reduced in VR group (mean difference –1.2, SD 2.9) vs control (mean difference –0.3, SD 1.7).
Mosso-Vasquez et al [17]	VR and analgesia	N/A	VNS	88% of patients reported a reduction in pain after VR therapy. Mean change in the Likert pain scale was 3.75. No descriptive statistics were provided.
				Change in pain scores (post procedure – pre procedure) was substantially correlated with change in respiratory rate ($R^2=0.26$). It was, however, minimally correlated with heart rate ($R^2=0.05$), mean arterial pressure ($R^2=0.09$), and SpO_2^m ($R^2=0.00$).
Mosso-Vasquez et al [34]	VR and analgesia	Mobile VR and analgesia	VAS	Overall, both head mounted display (presurgery: 6.06; postsurgery: 1.73) and mobile groups (presurgery 3.78; postsurgery 0.64) showed significant reductions ($P<.01$) in pain with VR.
				Head mounted display VR group experienced a significantly greater pain reduction from intra to postoperative states in comparison to the mobile VR group (–1.5 vs –0.07; $P=.02$).
Nilsson et al [19]	VR and analgesia	Analgesia alone	CAS; FAS[n]; FLACC	No significant difference in CAS, FAS, and FLACC scores between VR and non-VR groups ($P>.05$). No descriptive statistics were provided.
Piskorz et al [35]	VR and analgesia	Analgesia alone	VAS	The VR group (mean 15.16, SD 20.51) reported significantly lower ($P<.02$) pain intensity compared to that of the control group (37.05 SD 30.66). Pain intensity was 59% lower in the VR group than in the control with a large effect size (Cohen $d=0.86$).
Shoorab et al [36]	VR and analgesia	Analgesia alone	VNS	Statistically significant reduction in the pain scores were observed during episiotomy repair in the VR group using ANOVA (VR effect: $f=88.6$, $df=1$, $P<.01$).
				VR group had lower pain scores during several phases of the procedure in comparison to those of the non-VR group ($P<.0001$): during the repair of the hymen (VR: mean 9.0, SD 12.6; non-VR: mean 23.6, SD 19.8), skin (VR: mean 16.7, SD 16.5; non-VR: mean 39.3, SD 22.5), and after the repair (VR: mean 6.0, SD 12.8; non-VR: mean 25.2, SD 14).
Tashjian et al [37]	VR	Nature video	VNS	Pain reduction in the VR group (preintervention: mean 5.4, SD 2.6; postintervention: mean 4.1, SD 2.7) was greater (percentage reduction: 24% vs 12.2%, $P<.01$) than that in the control group within subjects (preintervention: mean 5.4, SD 2.6; postintervention: mean 4.8 SD 2.7).
				Higher number of responders in VR in comparison to control (≥0.5 SD drop in pain) (65% vs 40%, $P<.01$).
Walker et al [38]	VR and analgesia	Analgesia alone	VAS	No significant difference in pain scores ($P>.05$) between VR group and control—average pain (VR: 44 mm; control: 43 mm) and worst pain (VR: 66 mm; control: 59 mm)—during the procedure.

XSL·FO
RenderX

Study	Intervention	Standard care; control	Measurement tools	Results
Yun Hua et al [39]	VR and analgesia	Analgesia alone	FACES; VAS; FLACC	Significantly less pain reported in the VR group compared to the control group before, during and after the dressing change ($P<.05$). Also, significantly lower scores during dressing change in VR vs control ($P<.05$): FACES—VR: mean 2.42, SD 1.85; control: mean 4.19, SD 2.12)
				VAS—VR: mean 4.35, SD 2.64; control: mean 6.25, SD 2.84)
				FLACC—VR: mean 4.18, SD 2.97; control: mean 7.36, SD 3.47).

[a]VR: virtual reality.

[b]N/A: Not applicable.

[c]FACES: facial analysis scale (such as the Wong-Baker Faces Scale).

[d]ANOVA: analysis of variance.

[e]FLACC: Face, Legs, Activity, Cry, Consolability.

[f]VNS: verbal numerical scale.

[g]VAS: visual analog scale.

[h]CHEO: Children's Eastern Ontario Hospital Pain Scale.

[i]MANOVA: multivariate analysis of variance.

[j]NPS: numerical pain scale

[k]CAS: color analog scale

[l]GRS: graphical representation scale.

[m]SpO$_2$: oxygen saturation.

[n]FAS: facial affective scale

Effect on Anxiety

Anxiety or stress was measured as a primary outcome in 44% of studies (8/18) [25,28,30-33,35,38]. Within these, 50% (4/8) demonstrated a statistically significant reduction in anxiety; 37.5% (3/8) demonstrated a statistically significant reduction in anxiety utilizing a between-group design [28,32,35]; and 12.5% (1/8) showed a statistically significant reduction in anxiety by means of a within-group design [25]. These results and the tools utilized to measure them are detailed in Table 3.

Table 3. Summary measures of studies in review which measured forms of anxiety.

Study	Measurement tools	Results
Chad et al [25]	McMurtry children's fear scale	Significant reduction in fear detected by parent due to VR[a] (mean 2.18; P=.05).
		Insignificant reduction in fear reported by child due to VR (mean 2.57; P=.43).
Frey et al [28]	VNS[b] (scale of 0-10)	Anxiety was significantly decreased –1.5 (95% CI –0.8 to –2.3) in the VR condition compared to that in the control condition.
		Significant difference in anxiety found between VR group and non-VR group using ANOVA[c].
Gershon et al [30]	VAS[d] (scale of 0-100); CHEO[e] Pain Scale	From the CHEO pain scale measure, the VR group had significantly fewer behavioral markers in comparison to controls for anxiety (P<.05). No summary measures were provided in paper.
Glennon et al [31]	5-point Likert scale for anxiety	Participants in the VR group did not experience a statistically significant decrease in anxiety in comparison to that in controls (P>.05).
Gold et al [32]	VAS (scale of 0-10); FAS[f]	Significantly less anxiety (P<.05) was reported and observed in the VR group (mean 1.90, SD 2.2) compared to that in the control group (mean 2.48, SD 2.07).
McSherry et al [33]	VNS	Anxiety scores were not significantly reduced (P>.05) in VR group (mean difference –1.3, SD 4.4) vs control (mean difference –0.4, SD 2.7).
Piskorz et al [35]	VAS	The VR group (mean 11.16, SD 18.58) reported significantly lower stress levels (P<.01) compared to those in the control group (mean 41.89, SD 40.89).
		Stress levels were 73.4% lower in VR group against control with a large effect size (Cohen d= 0.993).
Walker et al [38]	VAS	No significant difference between intraprocedural anxiety levels. No descriptive statistics were provided.

[a]VR: virtual reality.

[b]VNS: verbal numerical scale.

[c]ANOVA: analysis of variance.

[d]VAS: visual analog scale.

[e]CHEO: Children's Eastern Ontario Hospital.

[f]FAS: facial affective scale.

Effect on Physiological Parameters

The effect of virtual reality on physiological indicators of pain was investigated in 39% of studies (7/18) [17,19,30,34,37-39]; however, the parameters investigated varied between studies, encompassing measures such as heart rate, respiratory rate, oxygen saturation, galvanic skin response, blood pressure, and mean arterial pressure.

Of these, 29% of studies (2/7) demonstrated a significantly reduced heart rate in children or adolescents undergoing virtual reality therapy compared with that of the control group [30,39]. A summary of the results is presented in Table 4.

Table 4. Summary of physiological measures and side effects of studies in this review.

Study	Outcome measures	Measurement tools	Results
Frey et al [28]	Nausea; side effects	Questionnaire	No adverse effects reported. No significant differences in occurrences of nausea between VR[a] and control.
Gershon et al [30]	Physiology	Heart rate	Significant reduction in physiological parameters (heart rate) observed in VR group vs non-VR group vs that of the control during procedure (96.3 vs 103.8 vs 110.3 beats per minute, $P<.05$).
Gold et al [32]	Side effects	Likert scale (scale 1-6)	5.2% (n=4) of patients reported nausea, and 8% reported simulator sickness.
Hoffman et al [18]	Nausea	GRS[b]	Nausea ratings were negligible.
Mosso-Vasquez et al [17]	Side effects; physiology	Questionnaire; heart rate; mean arterial pressure; respiration rate; SpO_2[c]	Change in pain scores (postprocedure – preprocedure) was minimally correlated with heart rate ($R^2=0.05$), mean arterial pressure ($R^2=0.09$), and SpO_2 ($R^2=0.00$). 37.3% (25/67) of patients had reduced heart rate after VR therapy. 52.2% (35/67) of patients had reduced mean arterial pressure after VR therapy. 64% (14/22) of patients had reduced respiratory rate after VR therapy. None of these data were tested for statistical significance. 4.5% experienced side effects.
Mosso-Vasquez et al [34]	Physiology	Blood pressure	No significant change in systolic or diastolic blood pressure with VR use.
Nilsson et al [19]	Physiology	Heart rate	No statistically significant difference in heart rate between VR and control group.
Tashjian et al [37]	Physiology; side effects	Questionnaire; blood pressure; heart rate	No adverse side effects reported. No statistically significant differences between systolic blood pressure, diastolic blood pressure, and heart rate pre- and post-VR ($P>.05$).
Walker et al [38]	Physiology; side effects	Questionnaire; heart rate; respiration rate; blood pressure; galvanic skin response	No significant difference between vital signs or galvanic skin response detected. No descriptive data provided. No side effects reported.
Yun Hua et al [39]	Physiology	Heart rate; SpO_2	Significantly lower heart rate was observed in the VR group compared to the control group (98.88 SD 11.57 vs 106.2 SD 11.45 beats per minute, $P<.05$). No difference in SpO_2.

[a]VR: virtual reality.

[b]GRS: graphic rating scale.

[c]SpO_2: oxygen saturation.

Side Effects

Of the eligible studies, 33% (6/18) assessed patients for side effects incurred from virtual reality therapy [17,18,28,32,37,38]. The main side effects that were screened were include nausea, vomiting, and vertigo. Overall, the prevalence of side-effects was low and ranged from 0% to 8%. This data is summarized in Table 4.

Discussion

Effect on Analgesia

The findings of this review illustrated that there was a significant reduction of pain related to virtual reality therapy utilization in 67% of the studies (12/18). Although acknowledging limitations in interpreting these findings (see limitations below), this evidence is corroborated by the findings of other high-quality studies [40-42], supporting the use of virtual reality therapy as a nonpharmacologic adjunct in facilitating analgesia within a clinical context. An effort was also made to critically appraise the studies which failed to demonstrate any significant differences in pain (Multimedia Appendix 3). The intention

here was to examine the studies for factors which may have contributed toward null findings and were considered to be limitations of the study by the authors themselves.

To the practicing clinician, these findings are of relevance as they suggest that virtual reality therapy can be considered as an inpatient adjunct for acute pain, particularly in the context of facilitating procedural analgesia (12/18). However, it is likely that this performance will vary by indication and the patient population to which it is being applied. This variability in performance, therefore, should prompt consideration toward pilot testing, as an initial step, for any specific clinical use in order to establish its appropriateness as a therapeutic modality.

To further elaborate, although the exact mechanisms behind how virtual reality facilitates analgesia are still unknown; there are several plausible theories which may explain its therapeutic effect. One school of thought suggests that virtual reality therapy enacts changes on a neurobiological level, and thereby, facilitates analgesia in a manner similar to a drug. Functional magnetic resonance imaging has been utilized to demonstrate this in experimental models [43]. During episodes of pain stimulus, areas of the neuroanatomic pain matrix (insula, anterior cingulate cortex, thalamus, primary and secondary somatosensory cortices) demonstrated increased levels of activity. When virtual reality therapy is administered to patients during these episodes, a reduction greater than 50% is observed in the activity of the pain matrix, which corresponds with a fall in patient-reported pain ratings [43]. Similarly, experimental models have also demonstrated that this analgesic effect of virtual reality therapy can be controlled in a dose-dependent fashion [44-46].

Alternatively, it has also been theorized that virtual reality enacts its functions on a psychological level through the distraction it provides. The Gate Control Theory [16] proposes that the amount of attention given to a painful stimulus affects the person's interpretation of it. In line with this, the Multiple Resource Theory [15] also suggests that humans have a finite capacity to provide attention toward and process pain. As such, it is plausible that by rerouting or drawing these mental faculties away from the noxious stimulus, through a mechanism such as virtual reality therapy, that this would successfully attenuate the perception of pain [42,47,48].

Effect on Anxiety

Within the brain, the limbic system and amygdala are implicated in mediating anxiety, and this is often experienced by patients prior to a medical procedure [49,50]. While the ability to be anxious is essential for survival, increased levels of anxiety in a clinical environment can lead to worsening perceptions of pain, decreased thresholds for pain, and less cooperative patients [20,21]. As such, a rationale does exist for controlling anxiety in the context of facilitating analgesia for patients.

The findings of the review with regard to anxiolysis were equivocal, with 50% (9/18) of the studies demonstrating a significant anxiolytic effect. A recent systematic review [51] demonstrated significantly reduced anxiety scores in individuals undergoing virtual reality therapy for treatment of anxiety disorders in comparison to those of controls. Similarly, there is

also some suggestion that virtual reality therapy generates positive emotions and improvements in mood which dampen preprocedural patient anxiety. Also, it is supposed that similar to its analgesic effects, anxiolytic properties occur as a result of the abovementioned psychological alterations [52-54].

There appears to be merit in further evaluating virtual reality therapy for its anxiolytic effect. Particularly, as the potential benefits of anxiolysis extend beyond the mitigation of procedural pain to include an improved patient experience [55].

Effect on Physiological Parameters

Several studies have attempted to utilize changes in physiological markers or autonomic arousal as surrogate marker of analgesic effect [56]. From a biological perspective, this is not surprising since acute pain activates the sympathoadrenal fight or flight response, which in turn produces autonomic arousal effects (ie, increased respiratory rate, heart rate, blood pressure, skin sweating—galvanic skin response) [22,56-58].

In this review, 39% of studies (7/18) explored the relationship between virtual reality therapy–facilitated analgesia and its effect on a variety of physiological parameters. Although our findings somewhat suggested that heart rate correlated with pain scores [30,39], it was not possible to consistently ascribe utility toward using autonomic arousal as a surrogate marker for analgesic effect. There were several reasons for this.

First, parameters utilized and investigated across studies appeared to be heterogeneous and inconsistent, making it difficult to draw firm conclusions. Next, there was also evidence to suggest that not all physiological markers respond similarly to pain stimulus and subsequent analgesia [17,19,30,34,37-39]. Additionally, it is also known that arousal induced by pain is not static as the participant may be able to influence it either consciously or subconsciously by utilizing their own coping strategies (ie, heightened respiratory rate can be consciously altered by slowing down one's breathing) [56,58,59]. This, therefore, will arguably impact the ability of physiological markers to be precise and consistent markers of pain. However, it is worth mentioning that the literature seems to suggest that both respiratory rate and galvanic skin response appear to be consistent markers of pain response, whereas cardiovascular changes appear to be less useful [56,58].

Considering these findings, it is safe to say that there is no firm evidence to suggest that virtual reality therapy can either affect autonomic arousal or demonstrate its analgesic properties through modulation of these parameters. Testing these parameters in a uniform and consistent manner, at least within a research context, is merited.

Side Effects

In this review, studies reported a low incidence (0%-8%) of adverse effects in participants utilizing virtual reality therapy; however, it should be reiterated that only 6 studies screened for side effects.

Some of the main side-effects associated with virtual reality were nausea, vomiting, eye strain, and dizziness; cumulatively referred to as cybersickness [60]. The most widely accepted theory explaining cybersickness relates to the Sensory Conflict

Theory. This refers to the discrepancy which occurs between the ocular and vestibular systems when the senses do not receive the usual sensory feedback that would be expected in such a scenario. This lack of synchronization is believed to cause cybersickness [61,62].

This is of relevance to the clinician using the technology for several reasons. For one, this alludes to a vulnerable population of patients who are susceptible to these side effects and who should be excluded from its use, such as patients with vestibular abnormalities, with seizure disorders, and who experience migraines or headaches [63]. Additionally, this alludes to a number of methods that can be considered to reduce the incidence of these effects during use of the technology. Although a discussion of these is beyond the scope of this review, the following articles provide adequate reference material [64-68].

Virtual Reality Technological Perspectives

Most virtual reality interactive hardware consists of a combination of a head mounted display, built-in biaural headphones for sound, and a trackpad or joystick for manipulation or navigation of the virtual environment, to provide the user with an immersive experience [69].

Head Mounted Display and User Control

In our review, all studies utilized a head mounted display for the administration of virtual reality to participants. These ranged from portable hardware, such as a helmet or piece of cardboard, to more sophisticated hardware systems, where participants were connected to an external processing unit.

A head mounted display displays content via 2 screens placed in front of the user's eyes which are stereoscopic in nature. The images displayed are angled to provide a variation in depth perception, which is interpreted by the brain as having 3D characteristics and features. In addition, the head mounted displays track user interaction in real time, which updates the virtual content that is reflected to the user simultaneously [70]. This can be either through tracking head orientation or position of the user's physical movements, as well as walking and jumping [69,71]. None of the studies in this review used systems capable of positional tracking. This is understandable, considering that clinical procedures, including those examined, typically necessitate controlled patient positioning.

Virtual Reality Content

In this study, 50% of the content was an active form of virtual reality, which entailed an element of interaction with the environment by the participant. In contrast, the remainder administered a passive form of virtual reality, where participants could only observe the content. This is worth mentioning as the available evidence suggests that the analgesia afforded by active virtual reality is significantly more than that offered by the passive form [72-74]. No study in our review, however, explicitly investigated this difference.

Limitations

The findings of this review should be interpreted considering the following limitations.

First, it is important to note that the results of this review could be influenced by publication bias. Particularly when considering that Fanelli et al [75] demonstrated that approximately 90% of literature in the fields of psychiatry, psychology, and clinical medicine report positive findings. As a result, the performance of virtual reality as an analgesic could be overestimated. Second, this limitation is further exacerbated by the use of a narrative approach employing descriptive statistics, as was the case for this review. Similarly, the search strategy was restricted to health databases given the interest in clinically relevant findings. However, in doing so, we neglected technical databases such as IEEE and ACM, which might have had further data of relevance. Unfortunately, this is an issue that plagues research in the field of medical technology as it attempts to incorporate the two very separate domains of health and technology. To bridge this, we have provided information in the discussion to supplement areas where more technical knowledge may have been required. Third, as the sample sizes of the studies included in this review were generally small and based on very specific inpatient populations, the generalizability of the findings may be limited. In addition to this, a large variety of measurement tools were implemented to quantify the outcome of pain. This unfortunately precludes meta-analysis of the data, which would have otherwise been useful to quantify accurate treatment effects.

Areas for Future Research

The systematic review highlighted the need for further large-scale prospective studies to be conducted in order to investigate the efficacy of virtual reality therapy in facilitating analgesia and anxiolysis. Additionally, this review also highlighted the need for investigators to screen patients for cybersickness-related side effects as part of their study design. Finally, it is also suggested that future clinical studies explore the differences between active and passive forms of virtual reality in facilitating analgesia.

Conclusion

This review sheds light on the efficacy of utilizing virtual reality therapy for the reduction of acute pain and procedural anxiety within an inpatient setting to hopefully offer a novel and practical perspective on the matter. Furthermore, it demonstrated a low incidence of adverse side effects in the populations being sampled. For clinical use, there appeared to be a preference for head mounted display to display virtual reality content. Although no differences between active and passive virtual reality were identified in this review, the literature appeared to suggest that active virtual reality would facilitate a higher level of analgesia in comparison to that facilitated by passive virtual reality [72-74]. Finally, although there was no evidence found to suggest an effect of virtual reality therapy on physiological parameters (autonomic arousal) as a surrogate biomarker of pain, this review also suggested merit in continuing this line of investigation in a rigorous and reproducible manner. It is hoped that this study serves to inform future trials to assess the efficacy of virtual reality in the treatment of acute pain.

Authors' Contributions

VS and RRW contributed equally to this paper. VS, FDSC, EMW, and BV conceived and designed the search. VS, RRW, JS, OP, AN, and SK analyzed the data. VS, RRW, JS, OP, and SK wrote the paper. VS, RRW, JS, OP, SK, FDSC, EMW, and BV interpreted the data and critically revised the manuscript. All authors reviewed and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General description of studies included in the review.
[DOCX File , 34 KB - jmir_v22i11e17980_app1.docx]

Multimedia Appendix 2
Technical specification of devices.
[DOCX File , 31 KB - jmir_v22i11e17980_app2.docx]

Multimedia Appendix 3
Critical appraisal of studies which failed to demonstrate a significant reduction in pain scores.
[DOCX File , 16 KB - jmir_v22i11e17980_app3.docx]

References

1. Merskey H, Bogduk N. Classification of Chronic Pain 2nd ed. Seattle: IASP Press; 1994. URL: https://s3.amazonaws.com/rdcms-iasp/files/production/public/Content/ContentFolders/Publications2/FreeBooks/Classification-of-Chronic-Pain.pdf [accessed 2020-10-23]
2. Federation of State Medical Boards of the United States, Inc. Model policy for the use of controlled substances for the treatment of pain. J Pain Palliat Care Pharmacother 2005;19(2):73-78. [Medline: 16061467]
3. Noble M, Treadwell J, Tregear S, Coates VH, Wiffen PJ, Akafomo C, et al. Long-term opioid management for chronic noncancer pain. Cochrane Database Syst Rev 2010 Jan 20(1):CD006605 [FREE Full text] [doi: 10.1002/14651858.CD006605.pub2] [Medline: 20091598]
4. Rosenblum A, Marsch LA, Joseph H, Portenoy RK. Opioids and the treatment of chronic pain: controversies, current status, and future directions. Exp Clin Psychopharmacol 2008 Oct;16(5):405-416 [FREE Full text] [doi: 10.1037/a0013628] [Medline: 18837637]
5. Gandhi K, Heitz JW, Viscusi ER. Challenges in acute pain management. Anesthesiol Clin 2011 Jun;29(2):291-309. [doi: 10.1016/j.anclin.2011.04.009] [Medline: 21620344]
6. Wardhan R, Chelly J. Recent advances in acute pain management: understanding the mechanisms of acute pain, the prescription of opioids, and the role of multimodal pain therapy. F1000Res 2017;6:2065 [FREE Full text] [doi: 10.12688/f1000research.12286.1] [Medline: 29225793]
7. Stewart WF, Ricci J, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. JAMA 2003 Nov 12;290(18):2443-2454. [doi: 10.1001/jama.290.18.2443] [Medline: 14612481]
8. Gregory J, McGowan L. An examination of the prevalence of acute pain for hospitalised adult patients: a systematic review. J Clin Nurs 2016 Mar;25(5-6):583-598. [doi: 10.1111/jocn.13094] [Medline: 26778249]
9. Won A, Bailey J, Bailenson J, Tataru C, Yoon I, Golianu B. Immersive virtual reality for pediatric pain. Children (Basel) 2017 Jun 23;4(7):52 [FREE Full text] [doi: 10.3390/children4070052] [Medline: 28644422]
10. Blondell R, Azadfard M, Wisniewski A. Pharmacologic therapy for acute pain. Am Fam Physician 2013 Jun 01;87(11):766-772 [FREE Full text] [Medline: 23939498]
11. Wiederhold BK, Soomro A, Riva G, Wiederhold MD. Future directions: advances and implications of virtual environments designed for pain management. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):414-422 [FREE Full text] [doi: 10.1089/cyber.2014.0197] [Medline: 24892206]
12. Chou R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, et al. Management of postoperative pain: a clinical practice guideline from the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain 2016 Feb;17(2):131-157. [doi: 10.1016/j.jpain.2015.12.008] [Medline: 26827847]
13. Trost Z, Zielke M, Guck A, Nowlin L, Zakhidov D, France CR, et al. The promise and challenge of virtual gaming technologies for chronic pain: the case of graded exposure for low back pain. Pain Manag 2015;5(3):197-206. [doi: 10.2217/pmt.15.6] [Medline: 25971643]

XSL•FO
RenderX

14. Hoffman HG, Doctor JN, Patterson DR, Carrougher GJ, Furness TA. Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. Pain 2000 Mar;85(1-2):305-309. [doi: 10.1016/s0304-3959(99)00275-4] [Medline: 10692634]

15. Wickens C. Processing Resources in Attention, Dual Task Performance, and Workload Assessment. USA: Defense Technical Information Center; 1981. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a102719.pdf [accessed 2020-10-23]

16. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965 Nov 19;150(3699):971-979. [doi: 10.1126/science.150.3699.971] [Medline: 5320816]

17. Mosso-Vázquez JL, Gao K, Wiederhold BK, Wiederhold MD. Virtual reality for pain management in cardiac surgery. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):371-378 [FREE Full text] [doi: 10.1089/cyber.2014.0198] [Medline: 24892200]

18. Hoffman HG, Patterson DR, Seibel E, Soltani M, Jewett-Leahy L, Sharar SR. Virtual reality pain control during burn wound debridement in the hydrotank. Clin J Pain 2008 May;24(4):299-304. [doi: 10.1097/AJP.0b013e318164d2cc] [Medline: 18427228]

19. Nilsson S, Finnström B, Kokinsky E, Enskär K. The use of virtual reality for needle-related procedural pain and distress in children and adolescents in a paediatric oncology unit. Eur J Oncol Nurs 2009 Apr;13(2):102-109. [doi: 10.1016/j.ejon.2009.01.003] [Medline: 19230769]

20. Andrasik F, Flor H, Turk DC. An expanded view of psychological aspects in head pain: the biopsychosocial model. Neurol Sci 2005 May;26 Suppl 2:s87-s91. [doi: 10.1007/s10072-005-0416-7] [Medline: 15926029]

21. Mosso J, Gorini A, De La Cerda G, Obrador T, Almazan A, Mosso D, et al. Virtual reality on mobile phones to reduce anxiety in outpatient surgery. Stud Health Technol Inform 2009;142:195-200. [Medline: 19377147]

22. Cowen R, Stasiowska MK, Laycock H, Bantel C. Assessing pain objectively: the use of physiological markers. Anaesthesia 2015 Jul;70(7):828-847 [FREE Full text] [doi: 10.1111/anae.13018] [Medline: 25772783]

23. Price D, Graham C, Parkin CG, Peyser TA. Are systematic reviews and meta-analyses appropriate tools for assessing evolving medical device technologies? J Diabetes Sci Technol 2015 Sep 29;10(2):439-446 [FREE Full text] [doi: 10.1177/1932296815607863] [Medline: 26420626]

24. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384. [doi: 10.1136/jech.52.6.377] [Medline: 9764259]

25. Chad R, Emaan S, Jillian O. Effect of virtual reality headset for pediatric fear and pain distraction during immunization. Pain Manag 2018 May;8(3):175-179. [doi: 10.2217/pmt-2017-0040] [Medline: 29722606]

26. Chan EA, Chung JW, Wong TK, Lien AS, Yang JY. Application of a virtual reality prototype for pain relief of pediatric burn in Taiwan. J Clin Nurs 2007 Apr;16(4):786-793. [doi: 10.1111/j.1365-2702.2006.01719.x] [Medline: 17402961]

27. Chau B, Chi B, Wilson T. Decreasing pediatric pain and agitation during botulinum toxin injections for spasticity with virtual reality: Lessons learned from clinical use. J Pediatr Rehabil Med 2018;11(3):199-204. [doi: 10.3233/PRM-180534] [Medline: 30223406]

28. Frey DP, Bauer ME, Bell CL, Low LK, Hassett AL, Cassidy RB, et al. Virtual reality analgesia in labor: the VRAIL pilot study-a preliminary randomized controlled trial suggesting benefit of immersive virtual reality analgesia in unmedicated laboring women. Anesth Analg 2019 Jun;128(6):e93-e96. [doi: 10.1213/ANE.0000000000003649] [Medline: 31094789]

29. Gerçeker GÖ, Binay ?, Bilsin E, Kahraman A, Yılmaz HB. Effects of virtual reality and external cold and vibration on pain in 7- to 12-year-old children during phlebotomy: a randomized controlled trial. J Perianesth Nurs 2018 Dec;33(6):981-989. [doi: 10.1016/j.jopan.2017.12.010] [Medline: 29559294]

30. Gershon J, Zimand E, Pickering M, Rothbaum BO, Hodges L. A pilot and feasibility study of virtual reality as a distraction for children with cancer. J Am Acad Child Adolesc Psychiatry 2004 Oct;43(10):1243-1249. [doi: 10.1097/01.chi.0000135621.23145.05] [Medline: 15381891]

31. Glennon C, McElroy S, Connelly L, Mische Lawson L, Bretches A, Gard A, et al. Use of virtual reality to distract from pain and anxiety. Oncol Nurs Forum 2018 Jul 02;45(4):545-552. [doi: 10.1188/18.ONF.545-552] [Medline: 29947355]

32. Gold J, Mahrer N. Is virtual reality ready for prime time in the medical space? a randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]

33. McSherry T, Atterbury M, Gartner S, Helmold E, Searles DM, Schulman C. Randomized, crossover study of immersive virtual reality to decrease opioid use during painful wound care procedures in adults. J Burn Care Res 2018 Feb 20;39(2):278-285. [doi: 10.1097/BCR.0000000000000589] [Medline: 28570305]

34. Mosso Vázquez JL, Mosso Lara D, Mosso Lara JL, Miller I, Wiederhold MD, Wiederhold BK. Pain distraction during ambulatory surgery: virtual reality and mobile devices. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):15-21. [doi: 10.1089/cyber.2017.0714] [Medline: 30256662]

35. Piskorz J, Czub M. Effectiveness of a virtual reality intervention to minimize pediatric stress and pain intensity during venipuncture. J Spec Pediatr Nurs 2018 Jan;23(1):1. [doi: 10.1111/jspn.12201] [Medline: 29155488]

XSL·FO
RenderX

36. JahaniShoorab N, Ebrahimzadeh Zagami S, Nahvi A, Mazluom SR, Golmakani N, Talebi M, et al. The effect of virtual reality on pain in primiparity women during episiotomy repair: a randomize clinical trial. Iran J Med Sci 2015 May;40(3):219-224 [FREE Full text] [Medline: 25999621]

37. Tashjian VC, Mosadeghi S, Howard AR, Lopez M, Dupuy T, Reid M, et al. Virtual reality for management of pain in hospitalized patients: results of a controlled trial. JMIR Ment Health 2017 Mar 29;4(1):e9 [FREE Full text] [doi: 10.2196/mental.7387] [Medline: 28356241]

38. Walker MR, Kallingal GJ, Musser JE, Folen R, Stetz MC, Clark JY. Treatment efficacy of virtual reality distraction in the reduction of pain and anxiety during cystoscopy. Mil Med 2014 Aug;179(8):891-896. [doi: 10.7205/MILMED-D-13-00343] [Medline: 25102532]

39. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. Pain Manag Nurs 2015 Oct;16(5):685-691. [doi: 10.1016/j.pmn.2015.03.001] [Medline: 25972074]

40. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag 2011 Mar;1(2):147-157 [FREE Full text] [doi: 10.2217/pmt.10.15] [Medline: 21779307]

41. Dascal J, Reid M, IsHak W, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]

42. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 01;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]

43. Hoffman HG, Richards TL, Bills AR, Van Oostrom T, Magula J, Seibel EJ, et al. Using FMRI to study the neural correlates of virtual reality analgesia. CNS Spectr 2006 Jan;11(1):45-51. [doi: 10.1017/s1092852900024202] [Medline: 16400255]

44. Loreto-Quijada D, Gutiérrez-Maldonado J, Nieto R, Gutiérrez-Martínez O, Ferrer-García M, Saldaña C, et al. Differential effects of two virtual reality interventions: distraction versus pain control. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):353-358. [doi: 10.1089/cyber.2014.0057] [Medline: 24892197]

45. Mahrer NE, Gold JI. The use of virtual reality for pain control: a review. Curr Pain Headache Rep 2009 Apr;13(2):100-109. [doi: 10.1007/s11916-009-0019-8] [Medline: 19272275]

46. Hoffman HG, Sharar SR, Coda B, Everett JJ, Ciol M, Richards T, et al. Manipulating presence influences the magnitude of virtual reality analgesia. Pain 2004 Sep;111(1-2):162-168. [doi: 10.1016/j.pain.2004.06.013] [Medline: 15327820]

47. Hoffman HG, Garcia-Palacios A, Kapa V, Beecher J, Sharar SR. Immersive virtual reality for reducing experimental ischemic pain. International Journal of Human-Computer Interaction 2003 Jun;15(3):469-486. [doi: 10.1207/S15327590IJHC1503_10]

48. McCaul KD, Malott JM. Distraction and coping with pain. Psychological Bulletin 1984;95(3):516-533. [doi: 10.1037/0033-2909.95.3.516]

49. Silverthorn D, Johnson B, Ober W, Ober C, Silverthorn A. Human Physiology: An Integrated Approach. London, UK: Pearson; 2016:2016.

50. Rosen JB, Schulkin J. From normal fear to pathological anxiety. Psychol Rev 1998 Apr;105(2):325-350. [doi: 10.1037/0033-295x.105.2.325] [Medline: 9577241]

51. Fodor LA, Cote CD, Cuijpers P, Szamoskozi S, David D, Cristea IA. The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: a meta-analysis. Sci Rep 2018 Jul 09;8(1):10323 [FREE Full text] [doi: 10.1038/s41598-018-28113-6] [Medline: 29985400]

52. Triberti S, Repetto C, Riva G. Psychological factors influencing the effectiveness of virtual reality-based analgesia: a systematic review. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):335-345. [doi: 10.1089/cyber.2014.0054] [Medline: 24892195]

53. Ganry L, Hersant B, Sidahmed-Mezi M, Dhonneur G, Meningaud J. Using virtual reality to control preoperative anxiety in ambulatory surgery patients: a pilot study in maxillofacial and plastic surgery. J Stomatol Oral Maxillofac Surg 2018 Sep;119(4):257-261. [doi: 10.1016/j.jormas.2017.12.010] [Medline: 29317347]

54. Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, et al. Evidence for striatal dopamine release during a video game. Nature 1998 May 21;393(6682):266-268. [doi: 10.1038/30498] [Medline: 9607763]

55. Smith V, Warty RR, Kashyap R, Neil P, Adriaans C, Nair A, et al. A randomised controlled trial to assess the feasibility of utilising virtual reality to facilitate analgesia during external cephalic version. Sci Rep 2020 Feb 21;10(1):3141 [FREE Full text] [doi: 10.1038/s41598-020-60040-3] [Medline: 32081989]

56. Kyle BN, McNeil DW. Autonomic arousal and experimentally induced pain: a critical review of the literature. Pain Res Manag 2014;19(3):159-167 [FREE Full text] [doi: 10.1155/2014/536859] [Medline: 24533429]

57. Terkelsen AJ, Andersen OK, Molgaard H, Hansen J, Jensen TS. Mental stress inhibits pain perception and heart rate variability but not a nociceptive withdrawal reflex. Acta Physiol Scand 2004 Apr;180(4):405-414. [doi: 10.1111/j.1365-201x.2004.01263.x]

58. Kyle BN, McNeil DW. Autonomic arousal and experimentally induced pain: a critical review of the literature. Pain Res Manag 2014;19(3):159-167 [FREE Full text] [doi: 10.1155/2014/536859] [Medline: 24533429]

59. Hohenschurz-Schmidt DJ, Calcagnini G, Dipasquale O, Jackson JB, Medina S, O'Daly O, et al. Linking pain sensation to the autonomic nervous system: the role of the anterior cingulate and periaqueductal gray resting-state networks. Front Neurosci 2020;14:147 [FREE Full text] [doi: 10.3389/fnins.2020.00147] [Medline: 33041747]

60. Almeida A, Rebelo F, Noriega P, Vilar E. Virtual reality self induced cybersickness: an exploratory study. In: Rebelo F, Soares M, editors. Advances in Ergonomics in Design. Advances in Intelligent Systems and Computing, vol 588. Cham, Switzerland: Springer Internatinal; 2018:A.

61. LaViola JJ. A discussion of cybersickness in virtual environments. SIGCHI Bull 2000 Jan 01;32(1):47-56. [doi: 10.1145/333329.333344]

62. Shafer D, Carbonara C, Korpi M. Modern virtual reality technology: cybersickness, sense of presence, and gender. Media Psychology Review 2017;11(2):1 [FREE Full text]

63. Gorini A, Riva G. Virtual reality in anxiety disorders: the past and the future. Expert Rev Neurother 2008 Feb;8(2):215-233. [doi: 10.1586/14737175.8.2.215] [Medline: 18271709]

64. Pausch R, Crea T, Conway M. A Literature Survey for Virtual Environments: Military Flight Simulator Visual Systems and Simulator Sickness. 1992 Jan;1(3):344-363. [doi: 10.1162/pres.1992.1.3.344]

65. Mourant RR, Thattacherry TR. Simulator sickness in a virtual environments driving simulator. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Nov 06;44(5):534-537. [doi: 10.1177/154193120004400513]

66. Lo W, So RH. Cybersickness in the presence of scene rotational movements along different axes. Applied Ergonomics 2001 Feb;32(1):1-14. [doi: 10.1016/s0003-6870(00)00059-4]

67. McCauley ME, Sharkey TJ. Cybersickness: perception of self-motion in virtual environments. 1992 Jan;1(3):311-318. [doi: 10.1162/pres.1992.1.3.311]

68. Rebenitsch L. Managing cybersickness in virtual reality. XRDS 2015 Sep 06;22(1):46-51. [doi: 10.1145/2810054]

69. Lee J, Ahn SC, Hwang J. A walking-in-place method for virtual reality using position and orientation tracking. Sensors (Basel) 2018 Aug 27;18(9):2832 [FREE Full text] [doi: 10.3390/s18092832] [Medline: 30150586]

70. de Franca A, Soares M. Review of virtual reality technology: an ergonomic approach and current challenges. In: de Franca A, Soares M, editors. Advances in Ergonomics in Design. AHFE 2017. Advances in Intelligent Systems and Computing, vol 588. Cham, Switzerland: Springer International; 2018:52-61.

71. Sherman W, Sherman WR, Craig A. Understanding Virtual Reality Interface, Application, and Design. San Francisco, CA: Elsevier Science; 2003.

72. Gutierrez-Martinez O, Gutierrez-Maldonado J, Cabas-Hoyos K, Loreto D. The illusion of presence influences VR distraction: effects on cold-pressor pain. Stud Health Technol Inform 2010;154:155-159. [Medline: 20543289]

73. Gutiérrez-Martínez O, Gutiérrez-Maldonado J, Loreto-Quijada D. Control over the virtual environment influences the presence and efficacy of a virtual reality intervention on pain. Stud Health Technol Inform 2011;167:111-115. [Medline: 21685651]

74. Gutierrez-Maldonado J, Gutierrez-Martinez O, Cabas-Hoyos K. Interactive and passive virtual reality distraction: effects on presence and pain intensity. Stud Health Technol Inform 2011;167:69-73. [Medline: 21685644]

75. Fanelli D. "Positive" results increase down the Hierarchy of the Sciences. PLoS One 2010 Apr 07;5(4):e10068 [FREE Full text] [doi: 10.1371/journal.pone.0010068] [Medline: 20383332]

Review

A Digitally Competent Health Workforce: Scoping Review of Educational Frameworks

Nuraini Nazeha[1], BA; Deepali Pavagadhi[1], MSc; Bhone Myint Kyaw[1], MBBS, MSc, PhD; Josip Car[1,2], MD, PhD, FRCPE, FFPH; Geronimo Jimenez[1,3], MA; Lorainne Tudor Car[2,4], MD, MSc, PhD

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[2]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[3]Department of Public Health and Primary Care, Leiden University Medical Center, Netherlands, Netherlands
[4]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Josip Car, MD, PhD, FRCPE, FFPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road
Singapore
Singapore
Phone: 65 69047005
Email: josip.car@ntu.edu.sg

Abstract

Background: Digital health technologies can be key to improving health outcomes, provided health care workers are adequately trained to use these technologies. There have been efforts to identify digital competencies for different health care worker groups; however, an overview of these efforts has yet to be consolidated and analyzed.

Objective: The review aims to identify and study existing digital health competency frameworks for health care workers and provide recommendations for future digital health training initiatives and framework development.

Methods: A literature search was performed to collate digital health competency frameworks published from 2000. A total of 6 databases including gray literature sources such as OpenGrey, ResearchGate, Google Scholar, Google, and websites of relevant associations were searched in November 2019. Screening and data extraction were performed in parallel by the reviewers. The included evidence is narratively described in terms of characteristics, evolution, and structural composition of frameworks. A thematic analysis was also performed to identify common themes across the included frameworks.

Results: In total, 30 frameworks were included in this review, a majority of which aimed at nurses, originated from high-income countries, were published since 2016, and were developed via literature reviews, followed by expert consultations. The thematic analysis uncovered 28 digital health competency domains across the included frameworks. The most prevalent domains pertained to basic information technology literacy, health information management, digital communication, ethical, legal, or regulatory requirements, and data privacy and security. The Health Information Technology Competencies framework was found to be the most comprehensive framework, as it presented 21 out of the 28 identified domains, had the highest number of competencies, and targeted a wide variety of health care workers.

Conclusions: Digital health training initiatives should focus on competencies relevant to a particular health care worker group, role, level of seniority, and setting. The findings from this review can inform and guide digital health training initiatives. The most prevalent competency domains identified represent essential interprofessional competencies to be incorporated into health care workers' training. Digital health frameworks should be regularly updated with novel digital health technologies, be applicable to low- and middle-income countries, and include overlooked health care worker groups such as allied health professionals.

XSL•FO
RenderX

Introduction

Background

Over the last three decades, there has been considerable interest in the use of digital health to enhance the quality, efficiency, and safety of health care [1,2]. Digital health and eHealth are often used interchangeably and broadly defined as "the use of information and communications technology in support of health and health-related fields" [3]. An analysis of various digital health definitions revealed three distinct yet overlapping uses such as monitoring, tracking, and informing health (eg, mobile devices, mobile sensors and wearables, apps, social media); enabling health communication among various stakeholders (eg, telehealth, virtual consultations); and storing, managing, and utilizing health data (eg, electronic medical records, electronic medication systems) [4]. Digital health tools have the potential to provide health care workers with a holistic view of patients' medical conditions through access to their health-related data and improved communication, regardless of distance and access [5]. Furthermore, the use of digital technology in health care could potentially reduce turnaround times, resource use, medication errors, and adverse drug events; increase the use of preventive care; and enable greater adherence to clinical guidelines [6-8].

Training and educating health care workers to be digitally competent is important for several reasons.

First, with the growing use of digital technology in health care, the roles and responsibilities of the health workforce are transforming in an unprecedented manner, intensifying the need for capacity building and continuous professional development. For example, a recent review commissioned by the United Kingdom Secretary of State for Health and Social Care (ie, the Topol Review) reported that within the next two decades, the majority of jobs in the National Health Service (NHS) will have a digital component [9]. Second, the importance and potential of remote care has been brought to light recently with the COVID-19 pandemic. Virtual consultation devices and electronic systems are indispensable tools used to diagnose and treat patients with potential COVID-19 infections as well as all other infections [10,11]. Third, even though the current and next generation of practitioners may be seen as "digital natives" [12], surveys of health care workers show that they would appreciate more training on digital technology [13,14]. Finally, improving digital literacy capabilities could lead to better adoption and implementation of digital services and technologies in health care settings [15]. Similarly, poor digital health literacy was found to be the most common barrier to the digital transformation of health care [16], and thus, the adoption of health technologies has been gradual in countries such as the United States [17], Europe [16], and Australia [18]. For the above-mentioned reasons, there is an increasing number of medical schools introducing digital health training into their curricula [12,19,20].

Such training programs should be guided by a clear framework of digital health competencies suited for different health care worker groups, settings, contexts, seniority, and role. Although there is an increasing number of individual digital health competency frameworks and reviews focusing on a specific health care worker role or setting [21-24], there is a need for consolidation, analysis, and a comprehensive overview of existing frameworks for all health care worker groups. This includes frameworks that are specific for and those that are relevant across different health professions, roles, or settings. Such an overview is important to inform increasingly interdisciplinary teams working in medicine and health care and corresponding future training initiatives, policy development, and research.

Objectives

The objective of this review is to identify and analyze the available digital health competency frameworks, regardless of health profession, role, or setting. This scoping review takes into consideration the heterogeneity and complexity of this field, and we aim to identify (1) the intended applications of digital health competency frameworks; (2) the methodologies employed; (3) the targeted audience in terms of health professions and settings; and (4) the type of competencies included in the frameworks. By doing so, we aim to provide an in-depth analysis of the existing frameworks as well as identify potential gaps and propose recommendations for the development of future frameworks and digital health training initiatives.

Methods

Study Design

We followed the guidelines by the Joanna Briggs Institute [25] in performing a literature review and guidelines by Tricco et al [26] in creating a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart. The protocol for this review was registered with the Open Science Framework [27]. In this review, we used the World Health Organization's definition of digital health as "the combination of e-health and m-health as well as emerging areas, such as the use of advanced computing sciences in big data, genomics and artificial intelligence" [3].

Eligibility Criteria

We included studies and reports focusing on the development and reporting of a digital health competency framework for health care workers. In this review, a competency framework is defined as a repository or a model that identifies, enlists, structures, and organizes competencies into meaningful categories and that has been developed via a systematic methodology or a relevant, established organization. This definition was developed *a priori* by referring to the existing definitions and descriptions, and looking at the common features of formerly identified digital health competency frameworks applicable to health care workers [28]. We included studies and reports on all health professions, including pre- and in-service, health care settings, and languages. Studies before January 2000 were excluded because digital health has evolved at a rapid pace, with substantial changes over the last two decades. The details of the inclusion and exclusion criteria are listed in Multimedia Appendix 1.

XSL•FO

RenderX

Search Strategy

The search strategy was developed collaboratively and iteratively by the reviewers with support from an experienced medical librarian and was guided by the following: (1) the sensitivity of the search strategy to relevant articles identified from previous manual searches and (2) the total number of relevant results in the first few pages of results in Medical Literature Analysis and Retrieval System Online (MEDLINE) and EMBASE (Excerpta Medica dataBASE). The final MEDLINE search strategy (Multimedia Appendix 2) was translated to other databases. Subsequently, the following 6 databases were searched on November 8, 2019: MEDLINE, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Education Resources Information Center (ERIC), PsycINFO, and the Cochrane Library. MEDLINE, EMBASE, ERIC, and PsycINFO were accessed via the Ovid platform, and CINAHL was accessed via EBSCOhost. We expected that there would be substantial unpublished work in this area, for which searches were performed using pertinent keywords (Multimedia Appendix 3) in OpenGrey, ResearchGate, and the first 10 pages of Google and Google Scholar. Websites of relevant professional associations (eg, International Medical Informatics Association [IMIA]), accreditation councils (eg, the US Accreditation Council for Graduate Medical Education [ACGME]), and key government websites (eg, Digital Health Canada, NHS Digital) were also searched (Multimedia Appendix 4).

Screening and Data Extraction

The reviewers screened the search results and assessed the full-text studies for inclusion. For the title and abstract screening, Covidence tool [29], a web-based software platform, and, for full-text screening, EndNote X8 were used. Subsequently, a data extraction form was used to record information from the selected full-text studies using Microsoft Excel. The form was developed to be in line with the research objectives and was piloted by reviewers on 3 studies. The form was further amended (Multimedia Appendix 5), and relevant data were extracted by reviewers. Each round of screening and data extraction process was performed by a pair of reviewers independently, and results were compared thereafter. Disagreements between reviewers were resolved by discussion, and where required, a third reviewer was engaged as an arbiter.

Data Synthesis

We analyzed the identified digital health competency frameworks in terms of their coverage of health professions, education level (eg, preservice or in-service), geographical applicability (ie, local or organizational, regional, national, international), health care settings (eg, acute care, emergency care, primary care), and other comparable features such as source and methodology. In frameworks that did not specify their methodology, an additional internet search was performed to retrieve information on the methods employed for framework development. Following this, we narratively synthesized and described framework characteristics, the evolution of frameworks over time, and the structural composition of competencies.

In addition, we performed a thematic analysis according to a list of steps proposed by Nowell et al [30] to understand the types of digital health competency categories presented in the frameworks. From the included frameworks, 2 reviewers independently classified competency categories into overarching domains. Studies that did not publish the full version of the framework were excluded from the thematic analysis, together with competency categories that were aimed at non–health care workers in a health care setting or irrelevant to digital competencies. Frameworks with ambiguous categories were excluded unless pertaining category descriptions or competencies were provided. Frameworks with competency themes or statements, with no distinct categories, were also included in the analysis by carefully allocating them to the identified domains. In some cases, where a single competency statement or theme encompassed several components of a competency, it was allocated to more than one relevant domain. After discussion and consensus, the reviewers derived a consolidated list of identified domains and their definitions and the prevalence of each identified domain across the included frameworks.

Results

Study Characteristics

The search generated 14,229 articles, of which 14,091 were from database searches and 138 from gray literature. After duplicates were removed and screening was completed, 33 articles were deemed eligible for inclusion. Of these, 27 articles presented individual frameworks. The remaining 6 articles presented preliminary findings followed by finalized versions of their frameworks (Staggers et al [31,32]; Egbert et al [33,34]; Hubner et al [35,36]), adding 3 more individual frameworks. As a result, a total of 30 frameworks are presented in this review (Figure 1), of which 16 were found through a gray literature search.

XSL•FO

RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart.
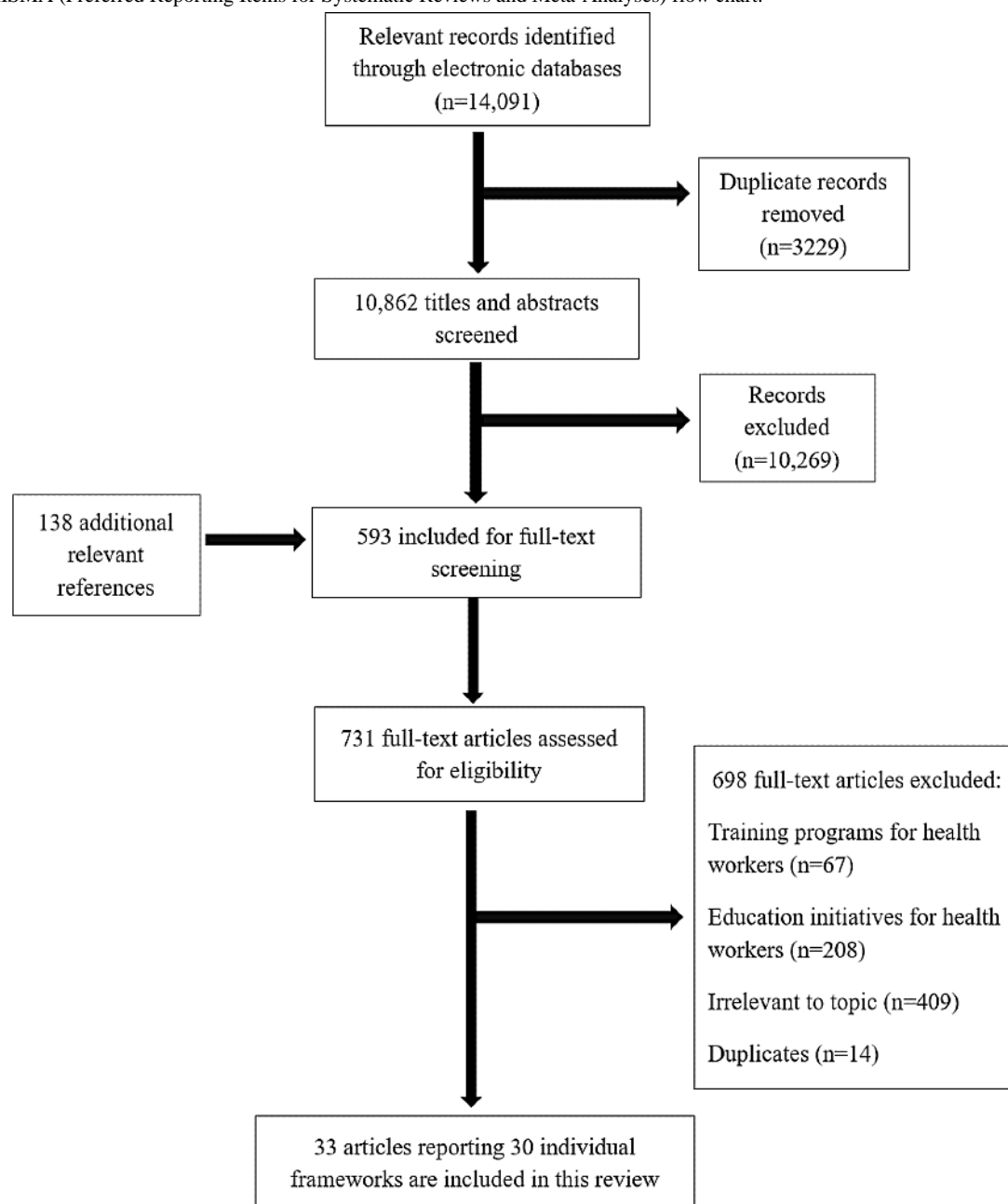
In terms of methodology, 14 frameworks employed literature review as an initial step, subsequently finalizing the frameworks through the use of expert consultations (n=5) [37-41], focus group discussions (n=3) [33,34,42,43], Delphi methodology (n=5) [31,32,44-47], or expert surveys [48]. In total, 11 other frameworks used one or a combination of methodologies (ie, Delphi, expert discussions, workshops, surveys) to reach consensus, largely by engaging various experts such as informaticists, health professionals, educators, and academics [21,22,36,49-56]. Of the remaining frameworks, 2 used only literature review to select relevant competencies [57,58], and 3 frameworks were built on the foundation of other published frameworks [59-61]. In addition, the frameworks were developed by a team of authors from a single university or institute (eg, University of Minnesota, School of Nursing) [44], by an international- or national-level organization (eg, IMIA, Australian Health Informatics Council [AHIEC]) [49,57], or

by means of a collaborative effort to produce frameworks such as Technology Informatics Guiding Education Reform (TIGER) [35,36,52] and Health Information Technology Competencies (HITCOMP) [41].

In terms of geographical relevance, 15 frameworks were country-specific [21,33,37,38,40,42,45,47,48,50,55-59], 1 was specific to the European Union region [61], 5 were applicable globally [35,36,41,49,52,54], and the remaining did not specify (n=9). In terms of health care settings, 4 were developed for remote care delivery [39,51,53,60], 1 framework each for hospitals [48], acute care [41], and homecare [50], while the remaining frameworks either were applicable to all health care settings (n=5) [42,45,56,57,59] or did not specify (n=18). In terms of health professions, 14 frameworks targeted nurses, 4 targeted doctors, of which one also included dentists, and 1 framework each targeted psychiatrists, dietitians, and public health professionals; 9 frameworks targeted health care workers

in general, of which 5 specified the inclusion of administrative, information technology (IT) support, and health informatics specialist roles [41,52,54,57,61], and 1 specified the inclusion of allied health professionals [39]. Among the nursing frameworks, 7 were meant for in-service nurses [21,31,32,42-44,46,48], 5 for preservice nursing students [33-37,53,58], and 2 for both [22,50]. Of the 4 medicine-focused frameworks, one focused on in-service doctors [40], another on preservice medical students [55], and 2 on both [59,60]. The framework characteristics and summary of the findings of the included studies are presented in Table 1 and Table 2, respectively. Additional details of the included studies are presented in Multimedia Appendix 6 [21,22,31-63].

Table 1. Characteristics of the 30 frameworks.

Characteristics	Values, n (%)
Publication dates	
2000-2010	6 (20)
2010-2019	24 (80)
Source	
Database	14 (47)
Gray literature	16 (53)
Geographical setting	
Country	15 (50)
Region	1 (3)
International	5 (17)
Not specified	9 (30)
Health care setting	
Acute care	1 (3)
Home care	1 (3)
Hospitals	1 (3)
Remote care	4 (13)
All health care settings	5 (17)
Not specified	18 (60)
Health care profession	
Nursing	14 (47)
Medicine	4 (13)
Allied health	1 (3)
Psychiatry	1 (3)
Public health professionals	1 (3)
Unspecified or applicable to multiple health professions	9 (30)
Educational level	
In-service	13 (43)
Preservice	9 (30)
Both	8 (27)

Table 2. Summary of findings of included studies.

Study ID	Context (country; health care setting)	Intended audience (profession; educational level)	Methodology	Organization of framework
Academy of Medical Royal Colleges (2011) [59]	Scotland; all	Doctors and dentists; preservice and in-service	Framework aligned with other competency frameworks	418 competencies divided into 20 domains, and further subdivided into outcomes (ie, knowledge, skills, behavior)
AHIEC[a] (2011) [57]	Australia; all	All HCPs[b] including admin and IT[c] support staff; preservice and in-service	Literature review drew on a wide range of major initiatives	45 competencies grouped into 3 categories and assigned a competency level ranging from 1 to 6
AFMC[d] in Partnership with Canada Health Infoway (2014) [55]	Canada; not specified	Medical students; preservice	Framework was based on contributions from experts	25 competencies classified according to physicians' roles and each competency is further divided into preclinical and clerkship milestones
Australian Nursing and Midwifery Federation (2015) [42]	Australia; all	Nurses and midwives (registered nurses, midwives, and enrolled nurses); in-service	• Literature review • Consensus via focus groups, on-line survey, and expert interviews	53 competencies grouped into 3 categories and 10 subcategories
Ayres (2012) [45]	United States; all	Registered dietitians, and dietetic technicians, registered and students; preservice and in-service	• Literature review • Delphi study	216 competencies grouped into 3 categories for each level of practice
Barakat (2013) [50]	The Netherlands; home care	Nurses; pre-service and in-service	A two-day collaborative workshop with academic experts	14 competencies organized into 5 themes
Brunner (2018) [47]	Australia; not specified	Health graduates; preservice	• Literature review • Focus group discussion • Delphi study	40 competencies organized into 4 domains and further divided into knowledge and performance cues
Chang (2011) [21]	Taiwan; not specified	Nurses; in-service	3 web-based Delphi rounds conducted with experts	318 competencies grouped into 3 categories for each level of practice
Collins (2017) [46]	Not specified	Nurse leaders; in-service	• Literature review • Delphi study (3 rounds) • Exploratory factor analysis for scale optimization and factor identification	74 competencies organized into 15 categories, and 15 of the most relevant competencies are ranked.
Crawford (2016) [51]	Not specified; remote care	Psychiatry residents; in-service	Expert panel and interviews with faculty and psychiatry residences	15 competencies sorted by professional roles, with examples
Curran (2003) [22]	Not specified	Nurse practitioners; preservice and in-service	Consensus with a team of nurse informaticists and nurse practitioner program directors	32 competencies grouped into 3 categories
Egbert (2016) [33]; Egbert (2019) [34]	Austria, Switzerland, Germany; not specified	Nurses; preservice	• Literature review • Expert survey • Focus group discussion and consensus	24 competency areas identified, and 5 of the most relevant areas ranked for 5 nursing roles
HITCOMP[e] (2019) [41]	International; acute care	All HCPs; preservice and in-service	• Literature review • Survey sent to international experts • Gap analysis • Expert consultation	1025 competencies organized into 33 competency areas, for 5 levels of practice across 5 domains
Hilty (2015) [60]	Not specified; remote care	Doctors, medical students; preservice and in-service	Competencies organized using the US ACGME framework, with input from the CanMEDS framework	Competencies listed for 8 main categories and subcategories, for each level of practice

XSL•FO
RenderX

Study ID	Context (country; health care setting)	Intended audience (profession; educational level)	Methodology	Organization of framework
Honey (2018) [58]	New Zealand; not specified	Registered nurses; preservice	• Curriculum mapping • Literature review	4 domains identified, and relevant subcategories and examples presented for each domain
Hubner (2016) [35]	International; not specified	Nurses; preservice	• Survey sent to international experts • A workshop was held to validate competencies	24 competency areas identified, and 6 of the most relevant areas ranked for 5 nursing roles
Hubner (2018) [36]	International; not specified	Nurses; preservice	• Survey sent to international experts • A workshop was held to validate competencies	• 24 competency areas identified, and sorted into 6 overarching domains • 10 of the most relevant areas ranked for 5 nursing roles
Hubner (2019) [52]	International; not specified	All HCPs; preservice	• Adapted Hubner (2016 and 2018)'s work [35,36] to include more HCP roles • Survey was sent to international experts	33 competency areas are identified, and the 10 most relevant competencies are ranked for each HCP role
Hwang (2008) [48]	Taiwan; hospital	Clinical nurses; in-service	• Literature review • Expert survey	49 competencies grouped into 3 main categories and subcategories
Jidkov (2019) [40]	United Kingdom; not specified	Doctors; in-service	• Literature review • Curricular content analysis • Expert consultation	20 competencies organized into 6 domains
Maheu (2018) [39]	Not specified; remote care	All HCPs including allied health professionals; Preservice and in-service	• Literature review • Expert consultation	7 domains identified, which are further broken down into 51 telebehavioral objectives, followed by 149 telebehavioral practices across 3 levels
Mantas (2010) [49]	International; not specified	All HCPs; in-service	Recommendations were discussed and refined by the IMIA[f] task force.	34 competencies organized into 3 BMHI[g] domains and each competency is determined if it is required by an IT user or a BMHI specialist according to 3 proficiency levels
Nagle (2014) [37]	Canada; not specified	Registered nurses; preservice	• Literature review • Consensus with experts through 3 rounds of feedback	19 competencies grouped into 3 domains.
NHS (2018) [56]	United Kingdom; all	All HCPs; in-service	Consultations with different stakeholders and workforce groups [62]	An overarching domain broken down into 5 domains, each with a set of description and competencies sorted according to 4 proficiency levels

XSL•FO
RenderX

Study ID	Context (country; health care setting)	Intended audience (profession; educational level)	Methodology	Organization of framework
Public Health Informatics Institute (2016) [38]	United States; not specified	Public health professionals; in-service	• Literature review • Expert consultation	8 categories identified, each with a set of competency statement and competencies
JASEHN (2018) [61]	Region; not specified	All HCPs including admin and IT support staff; in-service	• Literature review • Framework aligned with roles and competences as per the European eCompetence Framework • Use of framework descriptions to determine skill level for each role [63]	• Model 1: mission and main tasks described for 3 main profiles of health care workers (ie, health, nonhealth, and IT) • Model 2: 52 competencies grouped into 6 domains, where each competency has a description, and a set of associated knowledge and skill, according to 5 proficiency levels
Staggers (2001) [31]	Not specified	Nurses; in-service	• Literature review • Consensus with experts	304 competencies grouped into 3 categories for each level of practice
Staggers (2002) [32]	Not specified	Nurses; in-service	• Follow-up from Staggers (2001) [31] • 3 Delphi rounds were conducted	281 competencies grouped into 3 categories for each level of practice
Thye (2018) [54]	International; not specified	All HCPs including admin and IT support staff; preservice	• Mapped competency areas from Hubner (2018) [36] and HITCOMP [41] • Survey was sent to HCPs	33 competency areas identified, and the 10 most relevant interprofessional areas are listed
Trangenstein (2009) [43]	Not specified	Nurse scholars; in-service	• Literature review • Discussion and consensus	7 competency domains sorted according to nursing level of practice
Van Houwelingen (2016) [53]	Not specified; remote care	Nurses; preservice	• Survey with competencies was sent to participants • Delphi study (4 rounds)	52 competencies organized into 3 categories, where skills are further subdivided into 5 domains
Westra and Delaney (2008) [44]	Not specified	Nurse leaders; in-service	• Literature review • Delphi study	92 competency areas are grouped into 3 categories

[a]AHIEC: Australian Health Informatics Education Council.

[b]HCP: health care professional.

[c]IT: information technology.

[d]AFMC: Association of Faculties of Medicine of Canada.

[e]HITCOMP: Health Information Technology Competencies.

[f]IMIA: International Medical Informatics Association.

[g]BMHI: biomedical and health informatics.

Evolution of Frameworks

The digital health competency frameworks have drawn upon each other and have incrementally advanced the recommendations made in this area, as presented in Figures 2 and 3. The initial work of Staggers et al [31,32], which targeted nurses at 4 levels of practice, has been reproduced and adapted to suit different nursing roles (eg, nurse leaders, nurse practitioners) [22,44], health professions (eg, dietitians) [45], and even geographical settings (eg, Taiwan, Canada; Figure 2) [37,48]. In another instance, to propose competencies for nurse leaders, Collins et al [46] expanded and reorganized competency categories from Westra and Delaney [44], which initially drew

inspiration from the framework by Staggers et al [31,32]. The framework by Staggers et al [31,32] was updated 10 years later by Chang et al [21] with 42 new competencies.

Another commonly referenced framework is that by Egbert et al (Figure 3) [33,34]. This framework identified 24 core competency areas and conducted a survey with experts to rank the most relevant competency areas for nurses in Austria, Switzerland, and Germany. The same survey with 24 core competency areas was then sent to multiple countries to put forth international recommendations for nursing informatics, widely known as the TIGER framework [35,36]. The TIGER framework grouped the core competency areas into 6

overarching domains (ie, data, information and knowledge, information exchange and information sharing, ethics and legal issues, systems life cycle management, management in informatics, biostatistics, and medical technology) and ranked the most relevant competencies for nursing internationally.

Subsequently, the TIGER framework prompted its 2.0 version to include a wider spectrum of health care workers [52], and inspired the work by Thye et al [54] where interprofessional competencies were identified.

Figure 2. Digital health competency frameworks adapted from Staggers framework.
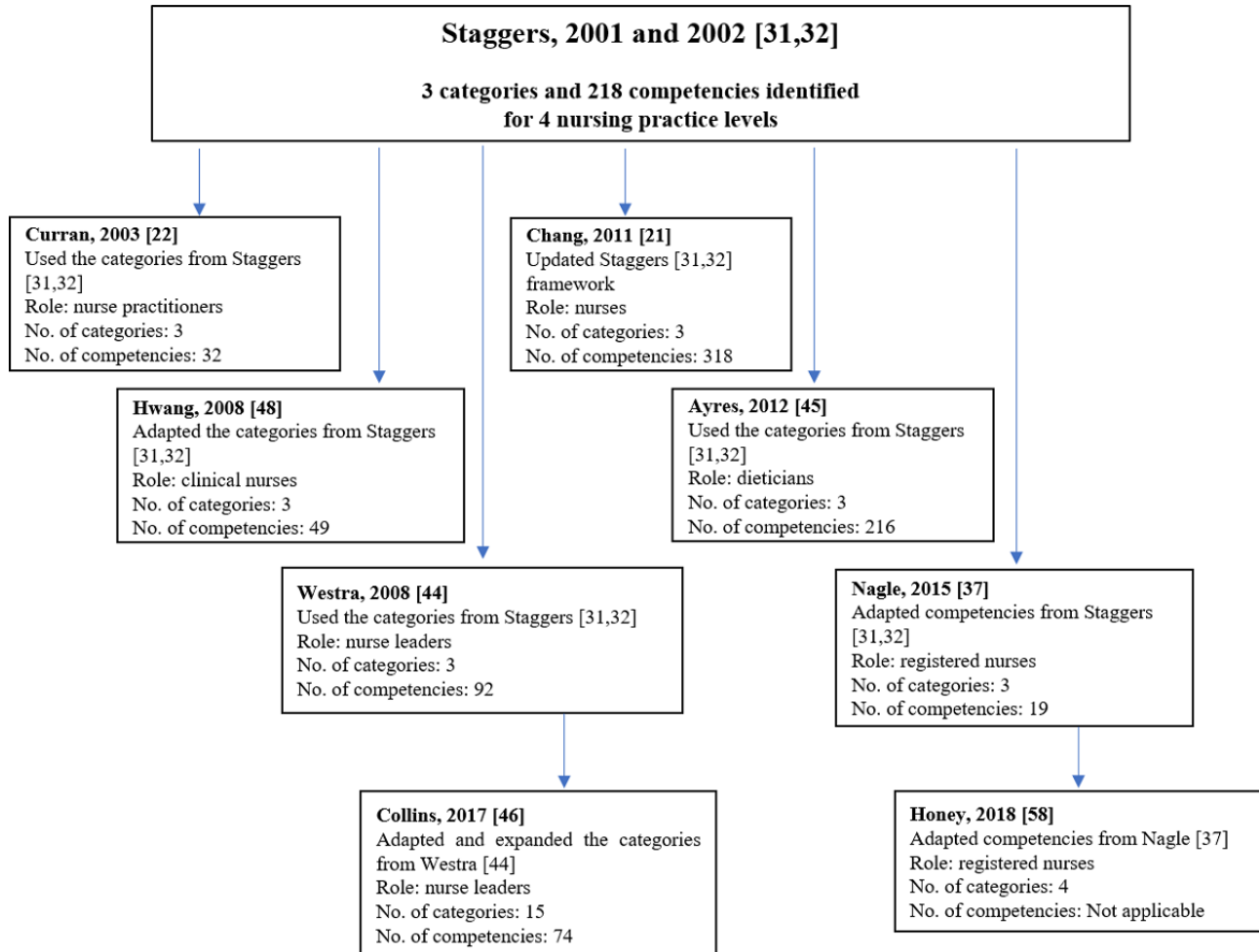
One framework that was accessible through an open-source internet-based database is the HITCOMP Tool and Repository [41]. HITCOMP was produced by the eHealth Workforce Development Workgroup as part of the EU*US eHealth Work Project. Its overall goal was to map, quantify, and project the needs of a digitally competent workforce [64]. This tool covers the digital competencies required in acute care settings for 5 broad roles of health care workers, similar to the TIGER framework version 2.0 (ie, direct patient care; administration; engineering or information, communication, and technology; informatics; and research or biomedicine) [52]. In comparison

with the other included frameworks, the HITCOMP framework has the highest number of competencies at 1025.

Other frameworks that inspired the development of subsequent frameworks include the IMIA [49], which was adapted by the AHIEC [57] to create national informatics standards for Australian health professionals, health informaticians, and specialists; the frameworks by Barakat et al [50] and the Academy of Medical Royal Colleges [59], which were adapted by Van Houwelingen et al [53] to develop a telehealth framework aimed at nurses; and the telepsychiatry framework by Hilty et al [60], which laid the foundation for telebehavioral health competencies by Maheu et al [39].

Figure 3. Development of recent digital health competency frameworks. HITCOMP: Health Information Technology Competencies; TIGER: Technology Informatics Guiding Education Reform.
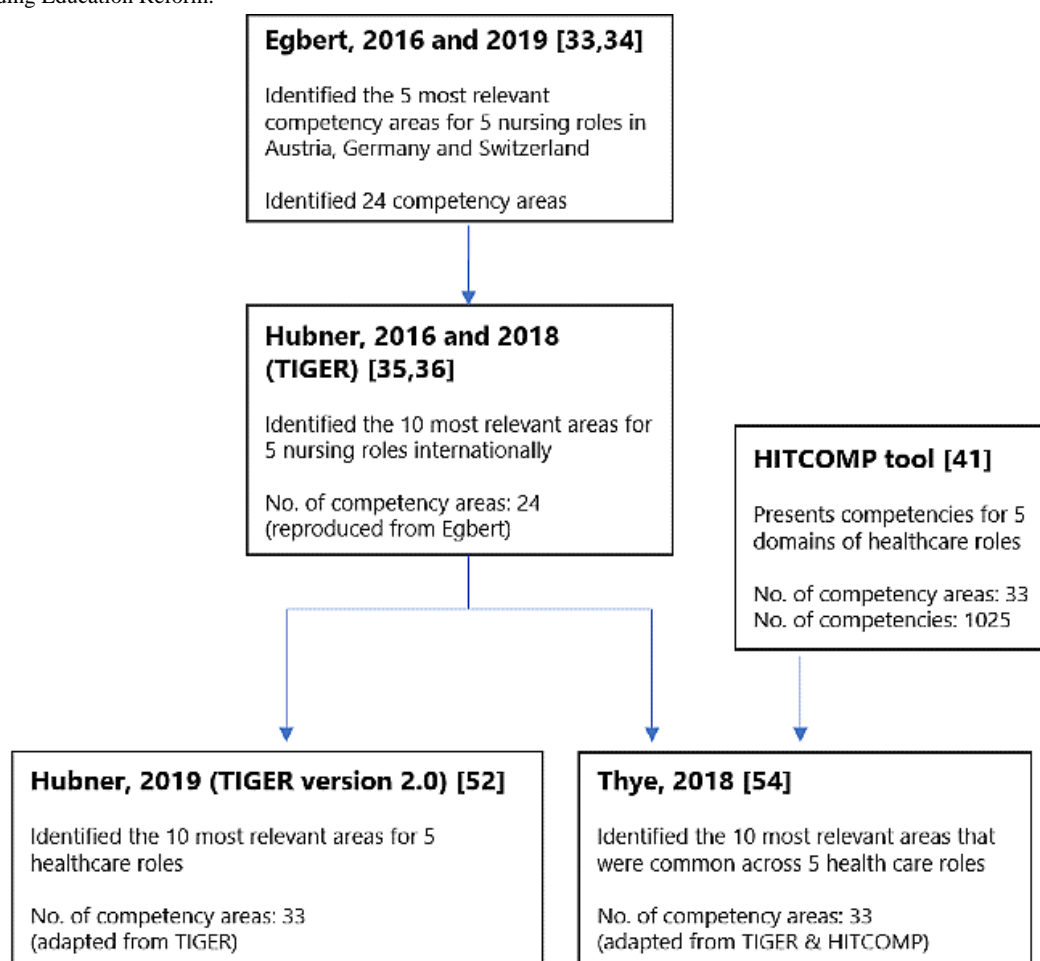
Digital Health Competencies and Categories

Across the included digital health competency frameworks, the number of competency categories, subcategories, and competencies ranged from 3 to 33, 10 to 39, and 15 to 1025, respectively. The wide range of reported competencies reflects the scale and specificity of the frameworks. For example, the framework with 15 competencies was focused on telepsychiatry training for psychiatry residents [51], while a framework with 318 competencies was intended for nurses at 4 different levels of practice [21]. Frameworks with multiple categories or subcategories had a larger number of competencies. For example, the eHealth competency framework by the Academy of Medical Royal Colleges presented 20 categories and included 418 competencies [59].

The primary objective of the frameworks was to guide the development of digital health curricula or training initiatives. Thus, 20 frameworks listed specific competencies (eg, knows how to use medical information systems for retrieval of patient data) [48], 6 frameworks cited case studies or provided examples for integration of competencies into curricula or training programs [33-36,45,51,52,58], and 3 frameworks ranked the most relevant competency areas, while the remaining 2 frameworks proposed only broad competency domains. Furthermore, in 12 frameworks, proficiency level in a digital health area or competency was presented according to the hierarchy of a profession or according to the depth of a skill that can be acquired by a health care worker role. For example, Chang et. al [21] presented competencies according to nursing staff seniority level (ie, a beginning nurse, an experienced nurse, an informatics specialist, an informatics innovator). Conversely, AHIEC assigned each competency a level of 1 to 6 (ie, 1: Remembering; 2: Understanding; 3: Applying; 4: Analyzing; 5: Evaluating; 6: Creating), according to the revised Taxonomy of Learning Domain objectives by Bloom, to indicate the depth of a skill that can be acquired by a health care worker role [57].

In addition, the structure of the included digital health competency frameworks varied. Competencies were organized either based on broad informatics categories (ie, computer skills, informatics knowledge, informatics skills) [21,22,31,32,42,44,45,48,53] or a combination of informatics and noninformatics categories (ie, information systems concepts, management concepts, ethical or legal concepts) [38,41,46,59-61] or based on health care worker roles (ie, communicator, collaborator, professional) [33-36,51,52,54,55]. Alternatively, competencies were sorted according to learning outcomes or statements [37,39,40,43,47,49,50,56-58]. For example, the national guideline for Canadian registered nurses sorted competencies according to 3 overarching statements: (1) uses relevant information and knowledge to support the delivery of evidence-informed patient or client care; (2) uses IT in accordance with professional and regulatory standards and

workplace policies; and (3) uses IT in the delivery of patient or client care [37].

Through thematic analysis, we were able to classify the majority of the competency categories presented in the frameworks into 28 domains. These domains are defined in Textbox 1, details of the analysis are provided in Multimedia Appendix 7 [45], and results of the analysis is presented in Multimedia Appendix 8. Competencies relating to the following domains were found to be prevalent in at least half of the included frameworks: informatics concepts and processes (22/30, 73%); health information and records management (19/30, 63%); communication (19/30, 63%); ethics, legal, or regulations (18/30, 60%); privacy and security (17/30, 57%); technical knowledge and support (15/30, 50%); and clinical care delivery (15/30, 50%). Conversely, competencies relating to medicines management (2/30, 7%) [41,59]; attitudes toward IT (4/30, 13%) [48,51,58,60]; IT advocacy (5/30, 17%) [21,22,51,55,60]; and public health (5/30, 17%) [41,49,52,54,55] were found to be uncommon in digital health competency frameworks. Of the 28 identified domains, 20 (71%) were present in at least one-third of the frameworks (Multimedia Appendix 8).

Textbox 1. Digital health competency domains identified from the included frameworks.

Administration and general management

- Use of administrative information technology (IT) applications to perform tasks and procedures, such as planning and delivery of services, business workflows, and human resource management

Analysis

- Use of IT systems to perform analysis of data, including data visualization, evaluation, and reporting

Attitudes toward IT

- Attitudes and cultural awareness toward the use of IT in patient care

Clinical care delivery

- Utilization of IT for the support of clinical care and practice, including use of assistive technologies and electronic test requesting

Communication

- Use of digital communications (eg, social media, email, etc) to enhance interpersonal skills and to aid in care delivery and decision-making process

Decision support

- Use of IT for clinical practice decision support

Documentation

- Use of IT for appropriate documentation tasks and processes, including knowledge of coding and terminologies

Education and training

- Use of IT in education and training, including e-Learning, mobile learning, and simulation

Ethics, legal, or regulations

- Knowledge of ethical, regulatory, compliance, and legal requirements relating to health IT

Financial management

- Knowledge of financial and account management relating to IT applications, including billing and fiscal aspects

Health information and records management

- Ability to access, collect, store, share, and manage digital health information; use of eHealth records; information and knowledge management

Health care quality and safety

- Ensuring or improving the quality and safety of health care services with the use of IT

Imaging

- Knowledge of biomedical imaging digital technologies

Informatics concepts and processes

- Knowledge and skills in computer basics, information systems, and general health IT use

Integration and interoperability

- Knowledge of integrated health IT applications, health information exchange, and interoperability standards, including coordination and collaboration aspects

IT advocacy

- Play an active role in promoting the use of IT in health care environments

Leadership and executive management

- Providing or enhancing executive leadership skills relating to the use of IT, including setting direction, strategic management, change management, stakeholder management, and governance

---

Medicines management

- Management of digital medication records and use of order entry

Patient access and engagement

- Promoting use of IT applications among patients and supporting or empowering patients for self-management, including patient access to patient health records

Privacy and security

- Ensuring that digital data and health information are protected and kept confidential by following privacy and security procedures

Project management

- Knowledge of project management software and associated terminologies

Public health

- Use of IT to inform public health strategies

Remote care

- Provision of care at a distance, including telehealth care, eHealth, mobile health, and related fields

Research

- Appropriate use of IT for research support and innovations

Risk management

- Managing IT-related risks

Systems implementation

- Knowledge and skills about IT systems development, management, and implementation

Technical knowledge and support

- Knowledge of technical aspects of IT systems, including software applications, testing, applied computer science, and IT maintenance and support capabilities

Discussion

Principal Findings

Of the 30 digital health competency frameworks, 14 solely targeted nursing staff. The frameworks predominantly originated from high-income countries and were developed based on literature reviews, followed by expert discussions or a Delphi approach. More than half of the included frameworks, especially those providing national-level recommendations, were from gray literature sources. Most frameworks were published between 2016 and 2019, highlighting the growing interest in digital health competencies in recent years.

The purpose of the retrieved digital health competency frameworks and the intended audience was clearly stated in most frameworks. The earliest frameworks and almost half of all the included frameworks were meant for nurses. This could be due to the significantly larger proportion of nurses in the health workforce [65]. Nurses play a crucial role in supporting health care environments by being a constant point of contact between patients and doctors; thus, there are various aspects of digital capabilities that are required of them (eg, using staff scheduling systems, extracting data from clinical systems, navigating decision support systems) [22]. These

nursing-focused frameworks have inspired subsequent works for other health care workers. For example, several competency areas subsumed under broad areas for nurses in the TIGER framework [36] were marked as standalone competency areas (ie, communication, legal issues, interoperability and integration, and life cycle management) for a wider spectrum of health care workers in version 2.0 of the TIGER framework [52]. Notably, only one framework was found for allied health professionals (ie, dietitians) [45], which highlights the perceived lack of interest in educating and training this group of workers in digital health. However, the Health Informatics Society of Australia, now known as the Australasian Institute of Digital Health, highlighted the need to focus on allied health care workers as their involvement is becoming increasingly important in decision making to improve patient care and health outcomes [66]. Furthermore, although there were frameworks for doctors and medical students collectively, only one framework was intended solely for undergraduate medical education. The framework was a national guideline for medical students in Canada briefly describing 25 eHealth competencies [55]. Medical practice relies heavily on communication, which is now achieved through various digital means; thus, the skills to utilize a range of digital technologies should be comprehensive and included in medical education [67]. Moreover, with the COVID-19 pandemic,

XSL·FO
RenderX

doctors are required to handle patient consultations digitally [11]. A 2018 survey conducted by European Medical Students' Association revealed that a majority of medical students rated their eHealth skills to be "poor" or "very poor" and desired for adequate digital health literacy [14].

Most of the included frameworks are useful for application in education or practice, mainly owing to the specificity of competencies, the organization of competencies according to proficiency levels or health care worker roles, and the illustration of case studies and examples to be applicable to various settings. On the other hand, frameworks by Trangenstein et al [43] and Jidkov et al [40] presented only broad competency themes, as it was believed that exhaustive lists of competencies could lead to poor adoption [40]. As the included digital health competency frameworks were heterogeneous in purpose, audience, and setting, it is challenging to determine a single framework as exemplary. Nevertheless, HITCOMP, which was developed via an iterative methodology, was found to be the most comprehensive framework, covering 21 out of the 28 identified competency domains, listing 1025 competencies, and targeting a wide audience of health care workers and medical specialties [41].

The key thrust of work in this area involves competencies related to informatics, followed by eHealth, telehealth or telebehavioral or telepsychiatry, digital capability, and health IT competencies. This distinction may be superficial, given that the definitions and terminologies seem to overlap across frameworks and the nomenclatural differences do not necessarily convey differences in competencies. For example, the interprofessional eHealth framework developed by Thye et al [54] utilized a range of informatics frameworks [33,41,49].

Although frameworks often drew upon each other, there were considerable variations among the identified competencies. The Egbert et al [33,34] framework identified the 5 most relevant competency areas (out of 24) for 5 nursing roles in Austria, Germany, and Switzerland. The TIGER framework [35,36] used the same competency areas and methodology as Egbert et al [33,34] and additionally reached out to experts worldwide for their inputs. The resulting relevant competency areas for a nurse in an IT management role, for example, varied considerably between both frameworks. For this role, competency areas, *risk management* and *project management*, ranked the top 5 internationally in the TIGER framework [35,36]; however, it was only relevant in 1 out of 3 countries in the Egbert et al [34] framework. In addition, eHealth, telematics, and telehealth, which were ranked as top 5 by Egbert et al [34] for 2 out of 3 countries (ie, Germany and Austria), did not make it to the top 10 list in the TIGER framework [35,36]. This suggests that the IT management role for nurses could be defined differently depending on each setting. Hence, a clear definition of the role is important to match the appropriate competency skills to a health care worker role.

In our thematic analysis of the competencies included in the retrieved frameworks, we identified 28 competency domains. The most prevalent domain relates to competencies aimed at providing knowledge on informatics concepts and processes. Examples of these include basic computer knowledge, information systems concepts, and principles of informatics, which are fundamental skills to health care workers intending to maximize the use of digital technologies. The other common domains included the ability of health care workers to manage data from health information systems and records and to be well-versed in digital communications. Most of the health-related data today exist in digital form; therefore, it is imperative for health care workers to be able to understand the purpose, basic structures, use, and storage of electronic health records (EHRs). In addition, as digital health entails new forms of communication (eg, virtual consultations, email, chatbots), it is imperative for health care workers to be able to relate information accurately yet efficiently, timely, and delicately to patients, colleagues, and other collaborating stakeholders [68]. Furthermore, the rise in the adoption and utilization of digital technologies has spurred new issues relating to the use of IT [69], which corresponds to the next two common domains of competencies (ie, ethics, legal, or regulations, and privacy and security). These domains stress the importance of health care workers' adherence to legal and regulatory requirements and to keep up to date with privacy and security policies pertaining to the appropriate use of digital technologies.

Some categories that were found in more recent frameworks, such as attitudes toward the use of IT, medication and prescription, IT advocacy, and public health, reflect the emerging trends in digital capabilities required by health care workers. For example, with the widespread adoption of EHRs and e-prescribing being a key functionality, it is imperative for physicians to be able to perform prescribing tasks efficiently and adapt to new features as systems continually evolve [70]. In addition, as digital technologies have an increasing role in the management of health of communities and populations, frameworks have also started to incorporate competencies related to public health [71]. Similarly, the acceptance of IT and its incorporation into everyday practice hinges on health care workers advocating for their use and being mindful of contextual factors and beliefs that would enable their use in different settings, such as high-income and low- and middle-income countries (LMICs). Other distinct domains such as leadership, administrative, managerial, and financial bring to attention that a digitally competent workforce should also be able to utilize technologies to oversee organizational-level aspects. It was also noted that competencies related to artificial intelligence, robotics, and social media, which are very relevant in current times, were not explicitly mentioned in the included frameworks [5,9,72]. One possible explanation could be that some of these are subsumed under broader categories, for example, competencies regarding the use of social media could be part of the *communication* category, or that these digital areas have yet to be covered in digital health competency frameworks.

From this synthesis of digital health competency frameworks, we would like to propose recommendations for the development of future frameworks (Textbox 2). First, an iterative methodology that includes literature review and consultations with local and international experts is ideal for a comprehensive framework. Second, it is encouraged for upcoming frameworks to explore competency areas that appeared in more recent frameworks, to cover upcoming digital health areas (ie, health

apps, artificial intelligence, autonomous decision-support systems), and to be open to future revisions to be up to date with technological developments. For example, HITCOMP is projected to continue mapping and aligning competencies with the curriculum and other major initiatives [64]. Finally, the lack of a comprehensive and international framework applicable to allied health professionals and LMICs warrants the development of frameworks that include these populations and settings. For example, only the TIGER framework includes case studies of LMICs, such as China, India, and Nigeria [73].

Textbox 2. Recommendations for the development of digital health competency frameworks.

Methodology:

- Literature review, followed by consultations or Delphi study with local experts, followed by engagement with international experts

Content:

- Explore new competency areas that appeared in recent frameworks (ie, attitudes and advocacy toward the use of information technology (IT), medication management, public health)

- Update competencies based on technological innovations and adoption and emerging evidence (ie, health apps, artificial intelligence, autonomous decision-support systems)

- Include essential interprofessional competencies (ie, informatics concepts and processes, health information management, communication, ethics, legal, or regulations, and privacy and security)

Target audience or setting to include:

- Allied health professionals

- Low- and middle-income countries

Application:

- Provide case studies of integration into curricula or training programs or examples of application in practice

---

The findings from this review can also inform and guide future training initiatives on digital health. When designing an educational or training program, it may not be possible to cover numerous competencies presented in a framework; rather, the program should focus on a specific set of competencies suitable for a particular group of health care workers, role, level of seniority, geographical, and health care setting. However, the identified competency domains prevalent in more than half of the included frameworks (ie, informatics concepts and processes, health information management, communication, ethics, legal, or regulations, and privacy and security) should be considered essential interprofessional competencies and thus should be incorporated into training and education efforts for any health care worker group. In addition, several digital competencies presented in the included frameworks may already be covered within the existing curriculum. For example, competencies within the *Analysis* and *Research* category may have been integrated within epidemiology training and evidence-based medicine education, respectively. Therefore, educators should consider integrating digital health training within existing parts of the curriculum and teaching it in an applied manner as much as possible. For example, the use of EHRs could be incorporated into the internal medicine curriculum.

Strengths and Limitations

To our knowledge, this is the first consolidation and analysis of existing digital health competency frameworks regardless of the role of health care workers. We performed a thorough search, including several databases and gray literature sources. Our analysis also provides a comprehensive overview of the types of competencies presented in digital health competency frameworks, which will aid in the training and education of health care workers to be digitally competent in relevant areas. Some weaknesses must be kept in mind when interpreting the findings of this review. Although a substantial number of frameworks from gray literature were included, some could have been missed, potentially national guidelines and standards from LMICs. In addition, when performing the thematic analysis, there were frameworks with vague competency categories and overlaps among some categories, leading to differences in opinions during the classification process. However, the 2 reviewers used their expertise to develop and clearly define the domains and allocate categories from frameworks into these domains, first independently and then through a consensus discussion, to reduce bias and classify as appropriately as possible. Although the reviewers aimed to make the classification process as transparent and reproducible as possible, it must be noted that there could be alternate ways of interpreting and classifying and that categorization may differ in the future with the new advances in digital health.

Conclusions

Of the 30 frameworks included in this scoping review, a majority target nurses, originate from high-income countries, and have been developed using an iterative approach. Our analysis of digital health competency frameworks can help inform the development of future digital health training programs for health care workers. Existing frameworks largely focus on the development of basic IT skills, proficiency in managing health-related information and digital communications, and awareness of ethical, legal, privacy, and security implications relating to IT. Future frameworks and training programs need to take into consideration the evolving nature of digital health and have to be able to incorporate upcoming digital trends, such

as artificial intelligence and robotics. There is also a need for frameworks focusing on LMICs, medical students, and allied health professionals.

Authors' Contributions

LC conceived the idea for the review and advised on the methodology. NN, DP, BK, and GJ were involved in the literature search, screening, and data extraction. NN and GJ were involved in the thematic analysis. NN analyzed and synthesized the data and drafted the manuscript. JC conceived the idea for the review and supervised the study. All authors reviewed the paper and provided critical inputs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion and exclusion criteria.
[[DOCX File , 50 KB](#) - [jmir_v22i11e22706_app1.docx](#)]

Multimedia Appendix 2
Search strategy for Medical Literature Analysis and Retrieval System Online (MEDLINE).
[[DOCX File , 48 KB](#) - [jmir_v22i11e22706_app2.docx](#)]

Multimedia Appendix 3
Keywords used for searching gray literature.
[[DOCX File , 52 KB](#) - [jmir_v22i11e22706_app3.docx](#)]

Multimedia Appendix 4
List of organizations whose websites were searched.
[[DOCX File , 48 KB](#) - [jmir_v22i11e22706_app4.docx](#)]

Multimedia Appendix 5
Data extraction form.
[[DOCX File , 48 KB](#) - [jmir_v22i11e22706_app5.docx](#)]

Multimedia Appendix 6
Details of included studies.
[[DOCX File , 68 KB](#) - [jmir_v22i11e22706_app6.docx](#)]

Multimedia Appendix 7
Allocation of competency categories or themes from included frameworks into overarching competency domains.
[[DOCX File , 76 KB](#) - [jmir_v22i11e22706_app7.docx](#)]

Multimedia Appendix 8
Competency domains prevalent across 30 digital health competency frameworks.
[[PNG File , 300 KB](#) - [jmir_v22i11e22706_app8.png](#)]

XSL•FO
RenderX

References

1. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]

2. Peckham D. Electronic patient records, past, present and future. Paediatr Respir Rev 2016 Aug(20 Suppl):8-11. [doi: 10.1016/j.prrv.2016.06.005] [Medline: 27394677]

3. WHO Guideline: Recommendations on Digital Interventions for Health System Strenghtening. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-10-08]

4. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is ehealth (6)? Development of a conceptual model for ehealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]

5. Global Diffusion of Ehealth: Making Universal Health Coverage Achievable. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2020-10-08]

6. Keasberry J, Scott IA, Sullivan C, Staib A, Ashby R. Going digital: a narrative overview of the clinical and organisational impacts of eHealth technologies in hospital practice. Aust Health Rev 2017 Dec;41(6):646-664. [doi: 10.1071/AH16233] [Medline: 28063462]

7. Nanah A, Bayoumi AB. The pros and cons of digital health communication tools in neurosurgery: a systematic review of literature. Neurosurg Rev 2020 Jun;43(3):835-846. [doi: 10.1007/s10143-018-1043-0] [Medline: 30334173]

8. Mohsin-Shaikh S, Furniss D, Blandford A, McLeod M, Ma T, Beykloo MY, et al. The impact of electronic prescribing systems on healthcare professionals' working practices in the hospital setting: a systematic review and narrative synthesis. BMC Health Serv Res 2019 Oct 22;19(1):742 [FREE Full text] [doi: 10.1186/s12913-019-4554-7] [Medline: 31640689]

9. The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future. The NHS Constitution. 2019. URL: https://topol.hee.nhs.uk/ [accessed 2020-10-08]

10. The Government of the Republic of Korea. Flattening the Curve on Covid 19: How Korea Responded to a Pandemic Using ICT. The Government of the Republic of Korea. 2020. URL: http://www.undp.org/content/seoul_policy_center/en/home/presscenter/articles/2019/flattening-the-curve-on-covid-19.html [accessed 2020-10-08]

11. Mian A, Khan S. Medical education during pandemics: a UK perspective. BMC Med 2020 Apr 9;18(1):100 [FREE Full text] [doi: 10.1186/s12916-020-01577-y] [Medline: 32268900]

12. Aungst TD, Patel R. Integrating digital health into the curriculum-considerations on the current landscape and future developments. J Med Educ Curric Dev 2020;7:2382120519901275. [doi: 10.1177/2382120519901275] [Medline: 32010795]

13. Stanford Medicine 2020 Health Trends Report: The Rise of the Data-Driven Physician. Stanford Medicine. 2020. URL: http://med.stanford.edu/content/dam/sm/school/documents/Health-Trends-Report/Stanford%20Medicine%20Health%20Trends%20Report%202020.pdf [accessed 2020-10-08]

14. Digital Health in the Medical Curriculum: Addressing the Needs of the Future Health Workforce. European Medical Students' Association Europe. URL: https://emsa-europe.eu/wp-content/uploads/2019/09/Digital-Health-in-the-Medical-Curriculum_-Addressing-the-Needs-of-the-Future-Health-Workforce.pdf [accessed 2020-10-08]

15. Improving Digital Literacy. Royal College of Nursing. 2017. URL: https://www.rcn.org.uk/professional-development/publications/pub-006129 [accessed 2020-10-08]

16. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]

17. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]

18. Evolution of eHealth in Australia: Achievements, Lessons, and Opportunities. Sydney: National E-Health Transition Authority Ltd. 2016. URL: https://tinyurl.com/yy3xg947

19. Mesko B, Győrffy Z, Kollár J. Digital literacy in the medical curriculum: a course with social media tools and gamification. JMIR Med Educ 2015 Oct 1;1(2):e6 [FREE Full text] [doi: 10.2196/mededu.4411] [Medline: 27731856]

20. Pathipati AS, Azad TD, Jethwani K. Telemedical education: training digital natives in telemedicine. J Med Internet Res 2016 Jul 12;18(7):e193 [FREE Full text] [doi: 10.2196/jmir.5534] [Medline: 27405323]

21. Chang J, Poynton MR, Gassert CA, Staggers N. Nursing informatics competencies required of nurses in Taiwan. Int J Med Inform 2011 May;80(5):332-340. [doi: 10.1016/j.ijmedinf.2011.01.011] [Medline: 21420902]

22. Curran CR. Informatics competencies for nurse practitioners. AACN Clin Issues 2003 Aug;14(3):320-330. [doi: 10.1097/00044067-200308000-00007] [Medline: 12909800]

23. Strudwick G, Nagle L, Kassam I, Pahwa M, Sequeira L. Informatics competencies for nurse leaders: a scoping review. J Nurs Adm 2019 Jun;49(6):323-330. [doi: 10.1097/NNA.0000000000000760] [Medline: 31135640]

24. Gonçalves LS, Wolff L, Staggers N, Peres A. Nursing informatics competencies: an analysis of the latest research. NI 2012 (2012) 2012;2012:127 [FREE Full text] [Medline: 24199069]

25. Peters M, Godfrey C, McInerney P, Baldini SC, Khalil H, Parker D. Scoping Reviews. In: Joanna Briggs Institute Reviewer's Manual. Australia: The Joanna Briggs Institute; 2017.

XSL•FO
RenderX

26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]

27. Pavagadhi D. Digital Health Competency Frameworks for Pre- and In-Service Health Workers from All Health Professions: A Scoping Review Protocol. OSF Registries. URL: https://doi.org/10.17605/OSF.IO/PXFRV

28. Milner RJ, Gusic ME, Thorndyke LE. Perspective: toward a competency framework for faculty. Acad Med 2011 Oct;86(10):1204-1210. [doi: 10.1097/ACM.0b013e31822bd524] [Medline: 21869668]

29. Covidence Systematic Review Software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2020-10-08]

30. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. Int J Quant Method 2017 Oct 2;16(1):160940691773384. [doi: 10.1177/1609406917733847]

31. Staggers N, Gassert CA, Curran C. Informatics competencies for nurses at four levels of practice. J Nurs Educ 2001 Oct;40(7):303-316. [Medline: 11596683]

32. Staggers N, Gassert CA, Curran C. A Delphi study to determine informatics competencies for nurses at four levels of practice. Nurs Res 2002;51(6):383-390. [doi: 10.1097/00006199-200211000-00006] [Medline: 12464758]

33. Egbert N, Thye J, Schulte G, Liebe J, Hackl W, Ammenwerth E, et al. An iterative methodology for developing national recommendations for nursing informatics curricula. Stud Health Technol Inform 2016;228:660-664. [Medline: 27577467]

34. Egbert N, Thye J, Hackl WO, Müller-Staub M, Ammenwerth E, Hübner U. Competencies for nursing in a digital world. Methodology, results, and use of the DACH-recommendations for nursing informatics core competency areas in Austria, Germany, and Switzerland. Inform Health Soc Care 2019;44(4):351-375. [doi: 10.1080/17538157.2018.1497635] [Medline: 30148411]

35. Hübner U, Shaw T, Thye J, Egbert N, Marin H, Ball M. Towards an international framework for recommendations of core competencies in nursing and inter-professional informatics: the tiger competency synthesis project. Stud Health Technol Inform 2016;228:655-659. [Medline: 27577466]

36. Hübner U, Shaw T, Thye J, Egbert N, Marin HD, Chang P, et al. Technology informatics guiding education reform - tiger. Methods Inf Med 2018 Jun;57(S 01):e30-e42 [FREE Full text] [doi: 10.3414/ME17-01-0155] [Medline: 29956297]

37. Nagle LM, Crosby K, Frisch N, Borycki E, Donelle L, Hannah K, et al. Developing entry-to-practice nursing informatics competencies for registered nurses. Stud Health Technol Inform 2014;201:356-363. [Medline: 24943567]

38. Applied Public Health Informatics Competency Model. Public Health Informatics Institute. 2016. URL: https://www.phii.org/resources/view/9462/applied-public-health-informatics-competency-model [accessed 2020-10-08]

39. Maheu MM, Drude KP, Hertlein KM, Lipschutz R, Wall K, Hilty DM. Correction to: an interprofessional framework for telebehavioral health competencies. J Technol Behav Sci 2018 Mar 5;3(2):108-140. [doi: 10.1007/s41347-018-0046-6]

40. Jidkov L, Alexander M, Bark P, Williams JG, Kay J, Taylor P, et al. Health informatics competencies in postgraduate medical education and training in the UK: a mixed methods study. BMJ Open 2019 Mar 30;9(3):e025460. [doi: 10.1136/bmjopen-2018-025460] [Medline: 30928942]

41. Health Information Technology Competencies. 2020. URL: http://hitcomp.org/competencies/ [accessed 2020-10-08]

42. National Informatics Standards for Nurses and Widwives. Australia: Australian Nursing and Midwifery Federation, Federal Office. 2015. URL: http://anmf.org.au/documents/National_Informatics_Standards_For_Nurses_And_Midwives.pdf [accessed 2020-10-08]

43. Trangenstein PA, Weiner EE, Gordon JS, McArthur D. Nursing informatics for future nurse scholars: lessons learned with the doctorate of nursing practice (DNP). Stud Health Technol Inform 2009;146:551-555. [Medline: 19592903]

44. Westra BL, Delaney CW, Delaney C. Informatics competencies for nursing and healthcare leaders. AMIA Annu Symp Proc 2008 Nov 6:804-808 [FREE Full text] [Medline: 18998803]

45. Ayres EJ, Greer-Carney JL, Fatzinger McShane PE, Miller A, Turner P. Nutrition informatics competencies across all levels of practice: a national Delphi study. J Acad Nutr Diet 2012 Dec;112(12):2042-2053 [FREE Full text] [doi: 10.1016/j.jand.2012.09.025] [Medline: 23174690]

46. Collins S, Yen P, Phillips A, Kennedy MK. Nursing informatics competency assessment for the nurse leader: the delphi study. J Nurs Adm 2017 Apr;47(4):212-218. [doi: 10.1097/NNA.0000000000000467] [Medline: 28333789]

47. Brunner M, McGregor D, Keep M, Janssen A, Spallek H, Quinn D, et al. An ehealth capabilities framework for graduates and health professionals: mixed-methods study. J Med Internet Res 2018 May 15;20(5):e10229 [FREE Full text] [doi: 10.2196/10229] [Medline: 29764794]

48. Hwang HG, Chen RF, Chang LH, Hsiao JL. A study of the informatics literacy of clinical nurses in Taiwan. Comput Inform Nurs 2008;26(5):290-299. [doi: 10.1097/01.NCN.0000304836.70133.d1] [Medline: 18769184]

49. Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, IMIA Recommendations on Education Task Force. Recommendations of the international medical informatics association (IMIA) on education in biomedical and health informatics. First revision. Methods Inf Med 2010 Jan 7;49(2):105-120. [doi: 10.3414/ME5119] [Medline: 20054502]

50. Barakat A, Woolrych RD, Sixsmith A, Kearns WD, Kort HS. Ehealth technology competencies for health professionals working in home care to support older adults to age in place: outcomes of a two-day collaborative workshop. Med 2 0 2013;2(2):e10 [FREE Full text] [doi: 10.2196/med20.2711] [Medline: 25075233]

51. Crawford A, Sunderji N, López J, Soklaridis S. Defining competencies for the practice of telepsychiatry through an assessment of resident learning needs. BMC Med Educ 2016 Jan 26;16:28 [FREE Full text] [doi: 10.1186/s12909-016-0529-0] [Medline: 26813286]

52. Hübner U, Thye J, Shaw T, Elias B, Egbert N, Saranto K, et al. Towards the tiger international framework for recommendations of core competencies in health informatics 2.0: extending the scope and the roles. Stud Health Technol Inform 2019 Aug 21;264:1218-1222. [doi: 10.3233/SHTI190420] [Medline: 31438119]

53. van Houwelingen CT, Moerman AH, Ettema RG, Kort HS, Ten Cate O. Competencies required for nursing telehealth activities: a Delphi-study. Nurse Educ Today 2016 Apr;39:50-62 [FREE Full text] [doi: 10.1016/j.nedt.2015.12.025] [Medline: 27006033]

54. Thye J, Shaw T, Hüsers J, Esdar M, Ball M, Babitsch B, et al. What are inter-professional ehealth competencies? Stud Health Technol Inform 2018;253:201-205. [Medline: 30147073]

55. eHealth Competencies for Undergraduate Medical Education. Canada: The Association of Faculties of Medicine of Canada. 2014. URL: https://www.infoway-inforoute.ca/en/component/edocman/3400-ehealth-competencies-for-undergraduate-medical-education/view-document?Itemid=101 [accessed 2020-10-08]

56. A HealthCare Digital Capabilities Framework. England: NHS. 2018. URL: https://www.hee.nhs.uk/sites/default/files/documents/Digital%20Literacy%20Capability%20Framework%202018.pdf [accessed 2020-10-08]

57. Health Informaticscope, Careers and Competencies. Australia: Australian Health Informatics Education Council. 2011. URL: https://www.ahiec.org.au/docs/AHIEC_HI_Scope_Careers_and_Competencies_V1-9.pdf [accessed 2020-10-08]

58. Honey M, Collins E, Britnell S. Guidelines: Informatics for Nurses Entering Practice. University of Auckland. 2018. URL: https://cdn.ymaws.com/www.hinz.org.nz/resource/resmgr/resources/Guidelines_Informatics_for_n.pdf [accessed 2020-10-08]

59. E Health Competency Framework - Defining the Role of the Expert Clinician. Scotland: Academy of Medical Royal Colleges. 2011. URL: https://www.physio-pedia.com/images/f/f0/EHealth_Competancy_Framework.pdf [accessed 2020-10-08]

60. Hilty DM, Crawford A, Teshima J, Chan S, Sunderji N, Yellowlees PM, et al. A framework for telepsychiatric training and e-health: competency-based education, evaluation and implications. Int Rev Psychiatry 2015;27(6):569-592. [doi: 10.3109/09540261.2015.1091292] [Medline: 26540642]

61. Recommendations on a Common Framework for Mapping Health Professionals' eHealth Competencies. Joint Action to Support the eHealth Network. 2018. URL: https://jasehn.eu/wordpress/wp-content/uploads/2018/07/JAseHN_D7.1.3_RECOMMENDATIONS_common_Framework_mapping_health_profession....pdf [accessed 2020-10-08]

62. Digital Literacy. Health Education England. URL: https://www.hee.nhs.uk/our-work/digital-literacy [accessed 2020-10-08]

63. Joint Action to support the eHealth Network JAseHN 677102 - Joint Actions. European Commission. URL: https://webgate.ec.europa.eu/chafea_pdb/health/projects/677102/workpackages [accessed 2020-09-10]

64. EU US eHealth Works to Improve Global Workforce Development. Technology Informatics Guiding Education Reform. 2017. URL: https://tutcris.tut.fi/portal/files/12978862/Eu_us_ehealth_workforce_development.pdf [accessed 2020-10-08]

65. Nursing and Midwifery. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/nursing-and-midwifery [accessed 2020-10-08]

66. Allied Health Professionals: the Untapped Potential in Digital Health. Health Informatics Society of Australia. 2019. URL: https://www.hisa.org.au/wp-content/uploads/2019/02/Allied-Health-Informatics-Position-Statement_draft.pdf [accessed 2020-10-08]

67. Wells KM. Social media in medical school education. Surgery 2011 Jul;150(1):2-4. [doi: 10.1016/j.surg.2011.05.023] [Medline: 21683853]

68. Griffiths F, Armoiry X, Atherton H, Bryce C, Buckle A, Cave J, et al. The role of digital communication in patient–clinician communication for NHS providers of specialist clinical services for young people [the Long-term conditions Young people Networked Communication (LYNC) study]: a mixed-methods study. Southampton (UK): NIHR Journals Library 2018. [Medline: 29485818]

69. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly 2018;148:w14571 [FREE Full text] [doi: 10.4414/smw.2018.14571] [Medline: 29376547]

70. Abramson EL, Patel V, Pfoh ER, Kaushal R. How physician perspectives on e-prescribing evolve over time. A case study following the transition between EHRS in an outpatient clinic. Appl Clin Inform 2016 Oct 26;7(4):994-1006 [FREE Full text] [doi: 10.4338/ACI-2016-04-RA-0069] [Medline: 27786335]

71. Gamache R, Kharrazi H, Weiner JP. Public and population health informatics: the bridging of big data to benefit communities. Yearb Med Inform 2018 Aug;27(1):199-206 [FREE Full text] [doi: 10.1055/s-0038-1667081] [Medline: 30157524]

72. Alraies MC, Raza S, Ryan J. Twitter as a new core competency for cardiologists. Circulation 2018 Sep 25;138(13):1287-1289. [doi: 10.1161/CIRCULATIONAHA.118.032999] [Medline: 30354418]

73. EU US eHealth Work Project H2020-SC1-HCO13-2016. Healthcare Information and Management Systems Society. 2016. URL: https://www.himss.org/euus-ehealth-work-projects-global-case-studies [accessed 2020-10-08]

Abbreviations

AHIEC: Australian Health Informatics Council
CINAHL: Cumulative Index to Nursing and Allied Health Literature
EHR: electronic health record
EMBASE: Excerpta Medica dataBASE
ERIC: Education Resources Information Centre
HITCOMP: Health Information Technology Competencies
IMIA: International Medical Informatics Association
IT: information technology
LMIC: low- and middle-income country
MEDLINE: Medical Literature Analysis and Retrieval System Online
NHS: National Health Services
TIGER: Technology Informatics Guiding Education Reform

XSL•FO
RenderX

Review

Use of Social Media to Promote Cancer Screening and Early Diagnosis: Scoping Review

Ruth Plackett[1], PhD; Aradhna Kaushal[2], PhD; Angelos P Kassianos[1], PhD; Aaron Cross[2], MSc; Douglas Lewins[3]; Jessica Sheringham[1], PhD; Jo Waller[4], PhD; Christian von Wagner[2], PhD

[1]Department of Applied Health Research, University College London, London, United Kingdom

[2]Research Department of Behavioral Science and Health, University College London, London, United Kingdom

[3]The Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis, Queen Mary University London, London, United Kingdom

[4]Cancer Prevention Group, School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom

Corresponding Author:
Ruth Plackett, PhD
Department of Applied Health Research, University College London
1-19 Torrington Place,
Bloomsbury
London, WC1E 7HB
United Kingdom
Phone: 44 20 3108 3242
Email: ruth.plackett.15@ucl.ac.uk

Abstract

Background: Social media is commonly used in public health interventions to promote cancer screening and early diagnosis, as it can rapidly deliver targeted public health messages to large numbers of people. However, there is currently little understanding of the breadth of social media interventions and evaluations, whether they are effective, and how they might improve outcomes.

Objective: This scoping review aimed to map the evidence for social media interventions to improve cancer screening and early diagnosis, including their impact on behavior change and how they facilitate behavior change.

Methods: Five databases and the grey literature were searched to identify qualitative and quantitative evaluations of social media interventions targeting cancer screening and early diagnosis. Two reviewers independently reviewed each abstract. Data extraction was carried out by one author and verified by a second author. Data on engagement was extracted using an adapted version of the key performance indicators and metrics related to social media use in health promotion. Insights, exposure, reach, and differing levels of engagement, including behavior change, were measured. The behavior change technique taxonomy was used to identify how interventions facilitated behavior change.

Results: Of the 23 publications and reports included, the majority (16/23, 70%) evaluated national cancer awareness campaigns (eg, breast cancer awareness month). Most interventions delivered information via Twitter (13/23, 57%), targeted breast cancer (12/23, 52%), and measured exposure, reach, and low- to medium-level user engagement, such as number of likes (9/23, 39%). There were fewer articles about colorectal and lung cancer than about breast and prostate cancer campaigns. One study found that interventions had less reach and engagement from ethnic minority groups. A small number of articles (5/23, 22%) suggested that some types of social media interventions might improve high-level engagement, such as intended and actual uptake of screening. Behavior change techniques, such as providing social support and emphasizing the consequences of cancer, were used to engage users. Many national campaigns delivered fundraising messages rather than actionable health messages.

Conclusions: The limited evidence suggests that social media interventions may improve cancer screening and early diagnosis. Use of evaluation frameworks for social media interventions could help researchers plan more robust evaluations that measure behavior change. We need a greater understanding of who engages with these interventions to know whether social media can be used to reduce some health inequalities in cancer screening and early diagnosis.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-033592

Introduction

Social media is becoming increasingly popular, estimated to be used by well over half of the world's population [1,2]. Given social media's potential for widespread public engagement, it is commonly incorporated into public health interventions [3,4]. The term "intervention" in this article refers to a spectrum of activities ranging from national awareness-raising public health campaigns, such as breast cancer awareness month (BCAM), to more targeted activities that use social media functions like discussion groups to deliver information to specific audiences. Social media interventions can influence behavior by targeting cognitive or emotional responses, generating discussions, and changing social norms [5]. Additionally, social media allows health messages to be disseminated rapidly, at low cost, to large numbers of people across large geographical areas [6-12].

Several national public health campaigns have used social media to try to improve the early diagnosis of cancer through raising awareness of cancer symptoms, encouraging help-seeking, and attempting to influence social norms around help-seeking [13,14]. However, there are concerns that the effects of these campaigns are short-lived, often only involve one-way communication, and do not benefit those in most need [4,5,15,16]. Social media can be used to target messages toward specific geographical regions or demographic groups, such as those who are known to have poor knowledge of cancer symptoms or encounter more barriers to accessing cancer screening [17,18]. Therefore, social media interventions may be more able to address health inequalities than traditional interventions [9,19,20]. However, little is known about the unintended effects of social media interventions and the possibility of spreading misinformation [9,21,22]. Additionally, social media cannot reach those with poor access to digital technology, who may also have the greatest need for public health information [23].

Despite increased use of social media interventions, there is little evidence about whether they improve cancer screening and early diagnosis [24]. There are no review publications that describe the variety of ways that social media specifically is used to promote cancer screening and early diagnosis, how these interventions might facilitate behavior change, and how this has been evaluated. Previous systematic reviews looking at a range of different media-based interventions for cancer screening included very few articles evaluating social networking sites like Facebook and have focused on specific research questions about impact and effectiveness [25,26]. However, use of social media interventions for public health is evolving rapidly and there is a need for a broader mapping of diverse studies to inform future development and evaluations. Therefore, a scoping review methodology was used to map the literature on the ways in which social media has been used to promote cancer screening and early diagnosis and how it was evaluated [27]. Specifically, we aimed to address the following research questions: (1) What are the characteristics of social media interventions that aim to promote cancer screening and early diagnosis?, (2) What are the mechanisms of change by which these interventions promote behavior change?, (3) What methodological approaches have been used to evaluate interventions?, (4) What are the outcomes used to measure the impact of interventions?, and (5) What are the key findings?

Methods

Overview

Social media interventions designed to promote cancer awareness and screening were identified using a scoping review guided by the methodology of Arksey and O'Malley [28] and expanded upon by Levac et al [29] and Peters et al [30]. We followed the process outlined in the published protocol and followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [31,32].

Search Strategy

An experienced research librarian helped to develop the search strategy, which included a combination of subject headings and keyword searches. We identified articles by searching five databases: MEDLINE, PsycINFO, Scopus, Web of Science, and CINAHL (Cumulative Index to Nursing and Allied Health Literature). Multimedia Appendix 1 shows the full search strategy used for MEDLINE, which was adapted for the other databases. Additional articles were identified by conducting internet searches for relevant published material and by hand-searching reference lists of included articles. We searched the grey literature for relevant reports not published in peer-reviewed journals. We also contacted the authors of conference abstracts that met inclusion criteria to see if we could include any unpublished results. Organizations and charities related to cancer screening were contacted via email for any relevant published research reports.

Inclusion and Exclusion Criteria

We considered peer-reviewed articles and non-peer-reviewed reports from the grey literature, as this was a scoping review that aimed to be inclusive and to explore the breadth of relevant research. The findings will help to inform future systematic reviews of the literature that will also assess the quality of the research. Additionally, non–peer-reviewed articles were included because many evaluations of social media interventions are not published in peer-reviewed journals. Social media is also a rapidly evolving area and current insights might not be captured in peer-reviewed literature, where there is typically a lag time between evaluation and publication. We excluded articles and reports not published in English, as we could not feasibly translate the results into English in a valid and reliable way. Articles and reports that discussed social media platforms, including Facebook, Twitter, Instagram, YouTube, Pinterest, and Snapchat, were eligible for inclusion. Articles relating to social media platforms popular in non–English-speaking

countries were not excluded, but as a consequence of limiting our search to articles written in English, the focus of this review is on platforms commonly used in English-speaking countries. We included articles published from 2004—as this was the advent of widespread social media use of these platforms—to June 2019 [33]. The reported findings from mass media campaigns were included if social media was the primary focus of the article. We included articles that discussed both qualitative and quantitative methods. We only included articles about interventions that directly targeted cancer screening and early diagnosis, where the primary message of the intervention focused on raising awareness of cancer symptoms, cancer screening, or promotion of help-seeking for potential cancer symptoms. Articles that reported findings solely on the human papillomavirus vaccine were excluded, as wider issues regarding attitudes toward vaccination and misinformation were considered out of the scope of this review. We excluded articles if the participants were cancer patients, survivors, or health professionals. Articles that evaluated static internet pages, such as blog posts, were also excluded from the review.

Article Screening and Data Extraction

All identified articles underwent two stages of screening: title/abstract screening and full-text screening. Three researchers (APK, AC, and SR) divided the articles and screened them against the inclusion criteria. Each article was independently reviewed by another researcher (AK). If the eligibility of any title/abstract was unclear, it was included in the full-text screening and any discrepancies were reviewed by an additional author and resolved in a consensus meeting. Interrater reliability for title/abstract screening was good (Cohen κ=0.69) [34]. Two researchers (RP and AK) piloted the data extraction approach

and three researchers (RP, APK, and AC) completed the data extraction. We extracted data using a Microsoft Access database to collect key information on article characteristics, details of the interventions, methodological approaches, outcome measures, and key findings. A second researcher (RP, APK, or AC) verified the data extractions, and discrepancies were resolved in regular meetings with the entire team.

Table 1 describes the outcomes extracted from the included articles. These were categorized using an adapted version of Neiger et al's [35] key performance indicators and metrics related to social media use in health promotion, which included insights, exposure, and reach, as well as low-, medium- and high-level engagement. Outcome measures capturing cancer knowledge and intention to attend cancer screening were not accounted for using the original framework but were categorized as high-level engagement, as greater knowledge and intentions are precursors to behavior change [36,37]. We also extracted information on the nature of information that was delivered and shared in interventions by users and developers, such as how many posts were related to health issues and fundraising. Additional data extraction was undertaken by one researcher (RP) to understand the mechanisms by which interventions might promote behavior change by using the behavior change technique (BCT) taxonomy developed by Michie et al [38]. BCTs were identified from the articles but also from campaign websites where possible. Following the scoping review methodology [28-30], we consulted with six people with experience developing and evaluating social media interventions for cancer screening, as well as a cancer patient, to validate our findings [39]. We asked them to comment on our preliminary findings, and their input helped to shape the narrative synthesis of the data.

Table 1. Descriptions and examples of outcomes captured during data charting.[a]

Data charting outcome	Description	Examples
Nature of information delivered/shared	Type of messages delivered by intervention or shared by users	Number of messages referring to cancer
Insights	User feedback	Users' opinions of information
Exposure	Views of social media content	Number of views of posts/tweets
Reach	Interaction with social media content and users' characteristics	Number of page likes and demographics of users
Low-level engagement	Agreement with the social media content	Number of likes of posts
Medium-level engagement	Users creating or sharing their own social media messages or sharing intervention messages on their own profiles	Number of posts/retweets
High-level engagement (behavior change)	Users' understanding of the messaging, intention to change their behavior, or actions taken offline related to the desired behavior change	Change in cancer screening attendance

[a]Insights, exposure, reach, and low-, medium-, and high-level engagement were measured using an adapted version of Neiger et al's [35] key performance indicators and metrics related to social media use in health promotion.

Results

Article Selection

Based on the initial search, 1029 articles were identified after duplicates were removed. We screened a total of 183 full-text

articles and included 23 of those articles in this review. Figure 1 outlines the selection process.

Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) flow diagram of the literature search and article selection process.
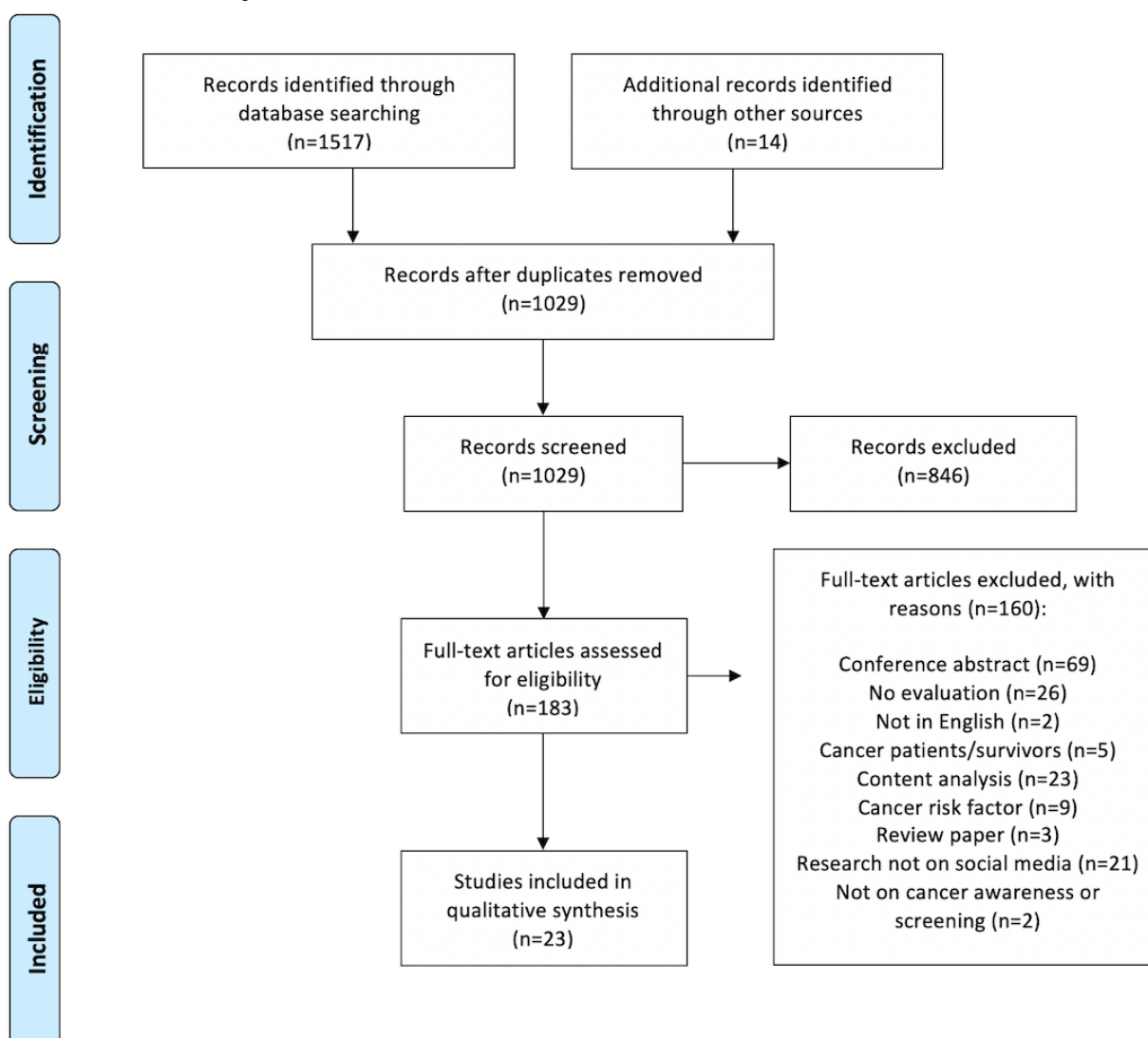
Characteristics of Articles

Multimedia Appendix 2 outlines the characteristics of each included article [40-62]. Only 2 of 23 (9%) articles were from the grey literature and were charity campaign reports; the rest of the articles were published in academic journals. First author institutions represented 10 countries, most commonly the United States (11/23, 48%), Canada (3/23, 13%), and the United Kingdom (2/23, 9%).

Characteristics of Social Media Interventions

Types of Interventions

Multimedia Appendix 2 describes the characteristics of each intervention. We found four different types of social media interventions present in the literature: (1) national cancer awareness month campaigns, (2) regional cancer awareness month campaigns, (3) targeted interventions, and (4) untargeted interventions. Most articles included in this review (16/23, 70%) evaluated national cancer awareness month campaigns, defined as national campaigns to improve cancer awareness and screening delivered to the general public; Multimedia Appendix

3 details an example campaign. Two articles (2/23, 9%) evaluated regional cancer awareness month campaigns. These were campaigns that took place at the same time as national campaigns but were delivered to a specific region [40,41]. Three articles (3/23, 13%) evaluated targeted interventions that used defined activities to deliver cancer awareness and screening information to specific groups and were not part of a cancer awareness month campaign. Multimedia Appendix 4 provides an example of a targeted intervention. Two articles (2/23, 9%) evaluated untargeted interventions that included cancer awareness and screening information delivered to the general public that were not part of an awareness month campaign. One untargeted intervention article (1/23, 4%) explored the impact of a celebrity Tweet about having cancer and the test that saved their life [42]. The other untargeted intervention delivered cancer information videos to the public [43].

Cancer Type

Most commonly, the articles reported interventions aimed at breast cancer (12/23, 52%), followed by prostate and testicular cancer (7/23, 30%), cervical cancer (4/23, 17%) [44-47], colorectal cancer (3/23, 13%) [42,48,49], generic cancer (no

specific cancer type; 3/23, 13%) [50-52], familial cancer (inherited tumors; 1/23, 4%), and lung cancer (1/23, 4%) [42,43]. Some articles evaluated interventions for more than one cancer type (7/23, 30%).

Intervention Source

Interventions were most commonly delivered by cancer charities (17/23, 74%), followed by public health government bodies (4/23, 17%) [40,44,47,53], regional health services (2/23, 9%) [43,54], a university (1/23, 4%) [43], and a celebrity (1/23, 4%) [42].

Social Media Platform

Interventions were most commonly delivered via Twitter (13/23, 57%), followed by Facebook (8/23, 35%), YouTube (3/23, 13%) [44,49,51], Instagram (2/23, 9%) [49,52], and Snapchat (1/23, 4%) [54]. Some articles evaluated interventions delivered via more than one platform (4/23, 17%).

Nature of Cancer Information Delivered and Shared in Interventions

Just over one-half of the articles (13/23, 57%) analyzed the messages that were delivered and shared by users and intervention developers. Several articles reported that most posts for national campaigns contained non-health messages and nonactionable messages [46,48,51,55-59]. During the 2013

Canadian Movember campaign, there were significantly more tweets on non-health topics, such as moustache growing (n=3549), than on health topics (n=673); only 0.6% (25/4222) of tweets analyzed were about cancer [56]. Furthermore, national campaigns heavily promoted online purchasing and fundraising to support cancer charities [55,57]. Bravo and Hoffman-Goetz [57] found that posts about fundraising and purchasing often did not mention cancer; only 2% (18/819) of fundraising tweets identified prostate or testicular cancer as the reason why they were fundraising. Additionally, for breast, cervical, prostate, and testicular cancer, gendered imagery and language were used to engage users [45,55,57]. In the Movember campaign, 9% (204/2400) of tweets analyzed used war metaphors, with users describing themselves as an "army," and the moustaches characterized as being "manly" [57].

Mechanisms of Behavior Change

BCTs Used in Interventions

No articles reported theories that informed the development of the intervention. Table 2 details the 10 BCTs that we tentatively identified as being present in the interventions [38]. All interventions targeted cancer screening behaviors such as improving attendance at cancer screening. This was achieved by providing information on cancer screening but also by raising awareness via fundraising activities.

Table 2. The number of interventions that used possible behavior change techniques (n=23) [38].

Behavior change techniques used in interventions	n (%)
Credible sources (eg, health professional, government, charity, celebrity)	23 (100)
Information about health consequences	23 (100)
Instruction on how to perform a behavior	22 (96)
Social support	17 (74)
Social comparison (eg, comparing a person's actions to the actions of others)	17 (74)
Information about others' approval	17 (74)
Goal-setting behavior	16 (70)
Social incentive (eg, providing a written reward only if a person performs the desired action)	12 (52)
Salience of consequences (eg, emphasizing the consequences of cancer)	2 (9)
Restructuring the physical environment (ie, changing the environment to facilitate the desired action)	1 (4)

The main way in which interventions appeared to influence behavior change was by providing information about the health consequences of cancer, often by providing links to cancer charity/public health websites to access cancer symptom and screening information [55,60]. The information delivered by interventions was all from credible sources, such as government bodies or cancer charities, but information from users may have been less credible. Information was made salient to users in posts that mentioned how users and members of their family were at risk of cancer (eg, "your children and family depend on it") [55,60]. Approximately three-quarters (17/23, 74%) of the articles reported interventions that used social support, social comparisons, and others' approval to encourage cancer screening. This was demonstrated by the sharing of personal stories and messages of support from the public, charities, organizations, cancer survivors, and celebrities [46,55,57]. One

article reported an intervention that restructured the environment by improving access to booking breast screening appointments through Facebook Messenger [61]. Interventions also influenced behavior change through goal setting and providing incentives. For example, many campaigns encouraged users to complete fundraising goals, such as growing a moustache for Movember to raise awareness [55-58]. They used social incentives like congratulating and thanking users on social media for taking part in fundraising activities to promote cancer screening [50,60].

Methodological Approaches Used to Evaluate Interventions

Multimedia Appendix 5 provides a breakdown of the methods used in each article. In most of the included articles (20/23, 87%), quantitative research designs were used; one-quarter

(5/20, 25%) of those were experimental research designs. A single-group posttest design was the most commonly used experimental design (3/5, 60%) [40,43,44]. Two articles (2/5, 40%) that used experimental designs used two-group pre- and posttest research designs [47,54]. Most (15/20, 75%) of the quantitative articles used observational research designs, where the researchers observed the impact of previously developed interventions. Just over one-half (8/15, 53%) of these were longitudinal studies. One-half (4/8, 50%) of the longitudinal studies measured outcomes over a period of 1 to 2 years [50-53]. Two (2/23, 8.7%) articles used qualitative research designs [55,57] and thematically analyzed social media messages. One (1/23, 4%) article used mixed methods by conducting a single-group pre- and posttest design and thematic analysis of social media messages [61].

Outcomes Used

Multimedia Appendix 5 shows the outcomes measured for each included article. Most (18/23, 78%) articles measured at least one of Neiger et al's [35] key performance indicators related to social media use in health promotion. Two (2/23, 9%) articles assessed insights by measuring user opinions [40,43]. Six (6/23, 26%) articles measured exposure, the most common indicator being impressions (the number of times a post was viewed; 4/6, 67%) [40,41,49,62]. Three (3/23, 13%) articles measured reach and the demographics of followers [40,49,52]. Five (5/23, 22%) articles measured low-level engagement by measuring the number of likes on a post [40,46,49,50,53]. Many (11/23, 48%) of the included articles measured medium-level engagement, with the most common indicator being the number of posts/tweets by users (8/11, 73%) [42,45,48,51,52,59,62]. Five articles (5/23, 22%) measured indicators of high-level engagement. Two (2/5, 40%) of these articles measured knowledge of cancer symptoms and screening [47,54] and one (1/5, 20%) article measured intention to get screened [40]. One (1/5, 20%) article measured participation in an offline advocacy event as a volunteer [41] and one (1/5, 20%) measured the number of people who attended screening [61].

Key Article Findings

Insights, Exposure, and Reach

Multimedia Appendix 5 presents the findings of each included article. Exposure varied by the type of intervention that was being evaluated. For example, one evaluation of a national campaign reported over 2 million Facebook impressions [49], and a regional campaign had 53,317 Facebook impressions [40]. Two (2/23, 9%) articles reported user insights and found social media was an acceptable way to deliver cancer awareness and screening information [40,43]. One (1/23, 4%) article found that 88% (43/49) of women surveyed indicated that they were neutral or agreed with seeing mammogram information on Facebook [40].

Three (3/23, 13%) articles suggested that reach varied by gender but only discussed the reach to women and men and no other gender identities. Content on YouTube may have reached more men than women, but Facebook content may have reached more women than men [40,43,49]. One (1/23, 4%) article found Facebook content had a wider reach to those aged 45 to 64 years

than to other age groups [40]. Another (1/23, 4%) article suggested that campaigns tended to reach more White users (93%) than African American (7%) or Asian or Hispanic users (0.6%) [49].

Low-Level and Medium-Level Engagement

One (1/23, 4%) article showed that users most commonly interacted with campaigns on Facebook by liking posts, followed by sharing content and commenting [50]. Two (2/23, 9%) articles showed that social media influencers and celebrities increased the number of likes due to their large number of followers [46,49]. Three (3/23, 13%) articles found that retweeting was significantly more likely if the tweet was posted by celebrities, organizations, someone with a high number of followers, or someone who frequently tweeted about the campaign [42,46,62]. Three (3/23, 13%) articles reported that posts with images were the most liked and were more likely to be retweeted than posts with just text [50,53,59].

Five (5/23, 22%) articles reported that engagement increased during the campaigns and decreased to baseline levels or below after the campaigns [41,50-52,62]. Two (2/23, 9%) articles found that health information–sharing tweets about cancer tended to rise during campaigns [42,48]. Two (2/23, 9%) articles reported that breast cancer campaigns had much more traffic on social media than other cancer campaigns, even on months that were dedicated to raising awareness of other cancers [42,51]. For example, even though the campaigns for prostate cancer awareness occurred in November, breast cancer received more mentions on Twitter in November than prostate cancer (284,015 posts versus 65,820 posts, respectively) [51]. Two (2/23, 9%) articles reported that colorectal cancer received the least attention on social media compared with breast, prostate, and cervical cancer [42,52]. Engagement with campaigns may vary by ethnicity, as one (1/23, 4%) article found White users consistently mentioned breast and prostate cancer more than other ethnicities [52].

High-Level Engagement

One (1/23, 4%) article reported that 9000 participants took part in an offline advocacy event that was part of a regional social media BCAM campaign [41]. Two (2/23, 9%) articles reported that targeted interventions improved knowledge of cancer symptoms and screening compared with a control [47,54]. One (1/23, 4%) article found that a regional mammography campaign improved intention to attend cancer screening; 82% of 49 women surveyed expressed an intent to get a mammogram in the next year [40]. One (1/23, 4%) targeted intervention reported an increase in the number of people who attended a breast screening appointment; attendance increased by an average of 12.9% across seven screening sites in North Midlands, United Kingdom (see Multimedia Appendix 4) [61].

Discussion

Principal Findings

Most studies of social media interventions have evaluated national cancer awareness month campaigns, using observational studies to measure exposure, reach, and low- to medium-level engagement with a campaign. A small number of studies

suggested that regional cancer awareness month campaigns and targeted interventions might improve cancer awareness, as well as screening intentions and uptake. There was evidence that exposure, reach, and engagement with the interventions varied by age, gender, and ethnicity of users, and also by cancer type.

This scoping review was the first to focus on social media evaluations of interventions to improve cancer screening and early diagnosis. It added to the literature by highlighting the limited number of robust evaluations that captured high-level engagement/behavior change, such as attendance at cancer screening. Evaluating high-level engagement is challenging because timely observational data on cancer screening attendance can be difficult to access and link with social media data [24]. Evaluation is also challenging because social media interventions are often designed without evaluation in mind [3]. During the consultation for this review, experts commented that campaigns are set up so quickly that there is not always time to consider evaluation. Improving cancer screening and early diagnosis is seen as a long-term goal that will take many years to realize, so immediate changes might not be expected or measured. A comprehensive evaluation framework that incorporates elements from behavior change theories and social media engagement frameworks could foster more robust evaluations that capture outcomes that demonstrate impact on behavior change and engagement [17,63,64]. However, as noted during the consultation for this review, there are further challenges to evaluation, such as the difficulty of demonstrating that a specific campaign caused a change in outcomes, as well as the limited time and resources of organizations to conduct evaluations.

This review was the first to use the BCT taxonomy to identify a variety of BCTs that social media interventions used to change health behaviors, including social support and providing information about health consequences [38]. One article reported restructuring the environment to provide better access to cancer screening [61]. As noted during our consultation, governance and data protection issues often limit the ability of health providers to use social media to improve access to care. These issues need to be addressed if interventions are to be tested or implemented on a larger scale. Many articles also reported that information delivered and shared during national cancer awareness month campaigns consisted of more non-health messages relating to fundraising than health messages relating to cancer. For many of these campaigns, the theory of change may be that fundraising messages increase cancer awareness, thereby increasing help-seeking or uptake of screening. Behavior change theory suggests that providing actionable health messages, such as information on cancer symptoms, could influence behavior change more directly than fundraising messages [12,38,63,65,66]. Future evaluations are needed to test our assumptions about how national campaigns might lead to behavior change and what messages would be most effective.

We identified a need for more social media interventions targeted toward colorectal and lung cancer, as most of the articles in this review were evaluations of social media interventions for breast and prostate cancer awareness and screening. Survival rates for lung and colorectal cancer are poorer than for breast and prostate cancer, which is partly because of poor uptake of cancer screening and delayed help-seeking that can lead to a delayed diagnosis [67]. Previous studies [68,69] have shown that there is more stigma around lung cancer than other cancers, and higher perceptions of cancer stigma are associated with delays in seeking medical care. Social media interventions could play a key role in changing social norms and stigma around help-seeking and screening for these cancers [70]. BCTs used by current breast cancer campaigns, such as sharing personal stories, could help to create social support and influence how people view these cancers, which in turn could encourage help-seeking behavior, increase screening uptake, and improve health outcomes.

This review also added to the literature by exploring to what extent inequalities in cancer screening and early diagnosis were measured and potentially addressed by social media interventions. We found some evidence that social media interventions have poorer reach and engagement with ethnic minorities, but there was no information on engagement with other minority groups [52]. Individuals from ethnic minorities might have less interaction with social media campaigns and not seek out cancer screening information because they have less access to cancer screening and higher cancer stigma [68,71-74]. The written and visual communication in social media interventions may also exclude ethnic minorities if the information is only available in English, presented in inaccessible language, or framed in a way that is unrelatable [74]. A lack of social media influencers or campaign role models that resonate with ethnic minorities may also make it less likely that they undertake a behavior, as suggested by social cognitive theory [75]. Future evaluations of social media interventions should measure inequalities in exposure, reach, and engagement and consider their success directly in relation to the groups that they were seeking to target. Targeting interventions toward those with a disproportionate disease burden could help to improve health inequalities seen in cancer screening and early diagnosis. However, social media interventions will have little impact on those who do not use social media, who may also be those in greatest need of information on cancer screening.

Some differences in the use of social media platforms by gender and age were found in this review [49,62]. Facebook content reached more women than men, and reached older-aged adults, and YouTube content reached more men than women [40,43,49]. This is consistent with recent data on social media use in the United Kingdom by age and gender [76]. This highlights the importance of identifying which platforms target users are currently more likely to engage with when designing social media interventions. However, as noted during our consultation, social media is a constantly changing landscape, so messaging needs to be continually updated and transferred to different platforms. Targeting messages also has potential risks, such as the threat to privacy and ethical issues, and it often requires payment and significant time and resources [17,24]. Additionally, data may not be available for particular target users; for example, we only found evidence to show how best to target those who identify as male and female. Previous research has shown that there are gender identity disparities in cancer screening, and trans and nonbinary individuals could benefit from more information regarding cancer screening and

early diagnosis [77,78]. Further research is needed to understand the most effective way of targeting social media interventions toward these individuals.

Limitations

We acknowledge that scoping reviews have several limitations, but a scoping review allowed us to gain a wide-ranging understanding of the role of social media in cancer screening. Research into social media is rapidly growing and this scoping review is a snapshot of evidence for social media interventions at a particular time [26]. Furthermore, as we only selected studies written in English that did not include information about some popular social networking sites in non–English-speaking countries, the findings of this review might not be generalizable. Many of the social media campaigns in this review were also part of multimedia campaigns. Therefore, it is difficult to know whether changes in engagement with social media or health behaviors were due to the social media element of those campaigns and how social media interacts with other aspects of the campaign. Additionally, the coding of BCTs in this review was dependent on reported content and online sources, so there was insufficient detail to identify all techniques used. The review was also dependent on what outcomes the evaluations chose to measure and report. There is currently no protocol for how to report evaluations of social media interventions, so there may be some degree of reporting bias in included articles. Future research would benefit from the development of a reporting protocol based on current frameworks for evaluating social media research [35].

Conclusions

This review found that most evaluations of social media interventions to improve cancer screening and early diagnosis did not report behavior change outcomes. The limited available evidence suggests some types of social media interventions may improve cancer awareness and intended and actual uptake of screening. Use of evaluation frameworks and reporting guidelines could help future researchers to plan robust evaluations of social media interventions that capture outcomes of behavior change and explore how these interventions work. Future evaluations could also measure who engaged with these interventions to assess whether social media interventions for cancer screening and early diagnosis can address some health inequalities. Interventions focusing on cancers that have received less social media attention, such as colorectal and lung cancer, could help to influence social norms around help-seeking and screening uptake for these cancers, which could improve health outcomes for patients.

Authors' Contributions

AK, APK, and AC were involved in the screening of titles and abstracts, as well as the full-text screening. RP, APK, and AC were responsible for data extraction. CVW, JW, JS, APK, and AK were involved in the design of the study. RP analyzed the data, and all authors contributed to the interpretation of the results and reviewed and approved the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[DOCX File , 13 KB - jmir_v22i11e21582_app1.docx]

Multimedia Appendix 2
Characteristics of included articles and interventions ordered by social media platform and target cancer of interventions.
[DOCX File , 18 KB - jmir_v22i11e21582_app2.docx]

Multimedia Appendix 3
Description of the Movember campaign.
[DOCX File , 17 KB - jmir_v22i11e21582_app3.docx]

Multimedia Appendix 4
Description of a targeted media intervention using Facebook.
[DOCX File , 14 KB - jmir_v22i11e21582_app4.docx]

Multimedia Appendix 5
Characteristics of the evaluations ordered by research design, including sampling, duration, outcome, and key findings.
[DOCX File , 26 KB - jmir_v22i11e21582_app5.docx]

References

1. Internet access - households and individuals, Great Britain. Office for National Statistics website. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018

2. Kemp S. Digital 2020: Global Digital Overview. DataReportal. 2020. URL: https://datareportal.com/reports/digital-2020-global-digital-overview

3. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]

4. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]

5. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]

6. Jackson D. Sprout Social. How to Go Viral Steps to Reach a Massive Audience. 2015. URL: https://sproutsocial.com/insights/how-to-go-viral/ [accessed 2020-06-19]

7. Smith BG, Kendall MC, Knighton D, Wright T. Rise of the Brand Ambassador: Social Stake, Corporate Social Responsibility and Influence among the Social Media Influencers. CMR 2018 Jul 30;03(01):6-29. [doi: 10.22522/cmr20180127]

8. Welch V, Petkovic J, Pardo PJ, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [Medline: 27077792]

9. Cavallo DN, Chou WS, McQueen A, Ramirez A, Riley WT. Cancer prevention and control interventions using social media: user-generated approaches. Cancer Epidemiol Biomarkers Prev 2014 Sep;23(9):1953-1956. [doi: 10.1158/1055-9965.EPI-14-0593] [Medline: 25103820]

10. Aiello AE, Renson A, Zivich PN. Social Media– and Internet-Based Disease Surveillance for Public Health. Annu. Rev. Public Health 2020 Apr 02;41(1):101-118. [doi: 10.1146/annurev-publhealth-040119-094402]

11. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]

12. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]

13. About Be Clear on Cancer. Cancer Research UK. 2014. URL: https://www.cancerresearchuk.org/health-professional/awareness-and-prevention/be-clear-on-cancer/about-be-clear-on-cancer

14. Breast Cancer Awareness Month. National Breast Cancer Foundation. 2020. URL: https://www.nationalbreastcancer.org/breast-cancer-awareness-month

15. Peacock O, Clayton S, Atkinson F, Tierney GM, Lund JN. 'Be Clear on Cancer': the impact of the UK National Bowel Cancer Awareness Campaign. Colorectal Dis 2013 Aug;15(8):963-967. [doi: 10.1111/codi.12220] [Medline: 23656572]

16. Bethune R, Marshall MJ, Mitchell SJ, Oppong C, Cartmel MT, Arumugam PJ, et al. Did the 'Be Clear on Bowel Cancer' public awareness campaign pilot result in a higher rate of cancer detection? Postgrad Med J 2013 Jul;89(1053):390-393. [doi: 10.1136/postgradmedj-2012-131014] [Medline: 23572594]

17. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018;1:47 [FREE Full text] [doi: 10.1038/s41746-018-0054-0] [Medline: 30854472]

18. Whyte-Smith H, Carpenter R, Koray E, Roe H. Technology Tracker Q2 2018. Ipsos. 2018. URL: https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-03/tech_tracker_q2_2018_final.pdf

19. Knoll J. Advertising in social media: a review of empirical evidence. International Journal of Advertising 2015 Mar 24;35(2):266-300. [doi: 10.1080/02650487.2015.1021898] [Medline: 000371744200007]

20. Chou WYS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]

21. Shah R. Fake news: A threat to curbing vaccine-preventable diseases in the UK. Nuffield Trust. 2019. URL: https://www.nuffieldtrust.org.uk/news-item/fake-news-a-threat-to-curbing-vaccine-preventable-diseases-in-the-uk

22. Moving the Needle: Promoting vaccination uptake across the life course. Royal Society for Public Health. 2019. URL: https://www.rsph.org.uk/static/uploaded/3b82db00-a7ef-494c-85451e78ce18a779.pdf

23.   Health and Digital: Reducing Inequalities, Improving Society. Tinder Foundation. 2016. URL: https://www.basw.co.uk/system/files/resources/basw_54514-4_0.pdf

24.   Pagoto S, Waring ME, Xu R. A Call for a Public Health Agenda for Social Media Research. J Med Internet Res 2019 Dec 19;21(12):e16661. [doi: 10.2196/16661] [Medline: 31855185]

25.   Schliemann D, Su TT, Paramasivam D, Treanor C, Dahlui M, Loh SY, et al. Effectiveness of Mass and Small Media Campaigns to Improve Cancer Awareness and Screening Rates in Asia: A Systematic Review. J Glob Oncol 2019 Apr;5:1-20 [FREE Full text] [doi: 10.1200/JGO.19.00011] [Medline: 30969807]

26.   Koskan A, Klasko L, Davis SN, Gwede CK, Wells KJ, Kumar A, et al. Use and taxonomy of social media in cancer-related research: a systematic review. Am J Public Health 2014 Jul;104(7):e20-e37 [FREE Full text] [doi: 10.2105/AJPH.2014.301980] [Medline: 24832403]

27.   Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]

28.   Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

29.   Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

30.   Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

31.   Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]

32.   Kaushal A, Kassianos AP, Sheringham J, Waller J, von Wagner C. Use of social media for cancer prevention and early diagnosis: scoping review protocol. BMJ Open 2020 Feb 25;10(2):e033592. [doi: 10.1136/bmjopen-2019-033592]

33.   The History of Social Media. Future Marketing. 2017. URL: https://www.future-marketing.co.uk/the-history-of-social-media/

34.   Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20(1):37-46. [doi: 10.1177/001316446002000104] [Medline: 5637250]

35.   Neiger BL, Thackeray R, Van WSA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]

36.   Fishbein M, Yzer MC. Using Theory to Design Effective Health Behavior Interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]

37.   Simon AE, Waller J, Robb K, Wardle J. Patient Delay in Presentation of Possible Cancer Symptoms: The Contribution of Knowledge and Attitudes in a Population Sample from the United Kingdom. 2010 Jul 26;19(9):2272-2277. [doi: 10.1158/1055-9965.epi-10-0219]

38.   Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

39.   Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews : a product of the ESRC methods programme (Version 1).: Lancaster, UK: University of Lancaster; 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2020-10-30]

40.   Klippert H, Schaper A. Using Facebook to communicate mammography messages to rural audiences. Public Health Nurs 2018 Oct 29;36(2):164-171. [doi: 10.1111/phn.12556]

41.   Salako O, Roberts AA, Isibor VI, Babatunde O, Fatiregun O, Nwogu CN. Innovative Breast Cancer Awareness and Advocacy Campaign. JGO 2017 Apr;3(2):169-176. [doi: 10.1200/jgo.2016.003509]

42.   Vos SC, Sutton J, Gibson CB, Butts CT. Celebrity Cancer on Twitter: Mapping a Novel Opportunity for Cancer Prevention. Cancer Control 2019 Feb 28;26(1):107327481982582. [doi: 10.1177/1073274819825826]

43.   Jones GE, Singletary JH, Cashmore A, Jain V, Abhulimhen J, Chauhan J, et al. Developing and assessing the utility of a You-Tube based clinical genetics video channel for families affected by inherited tumours. Familial Cancer 2016 Jan 11;15(2):351-355. [doi: 10.1007/s10689-016-9866-8]

44.   Cooper CP, Gelb CA, Chu J. Gynecologic Cancer Information on YouTube: Will Women Watch Advertisements to Learn More? J Cancer Educ 2016 Sep;31(3):602-604 [FREE Full text] [doi: 10.1007/s13187-015-0830-3] [Medline: 25877466]

45.   Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising Awareness About Cervical Cancer Using Twitter: Content Analysis of the 2015 #SmearForSmear Campaign. J Med Internet Res 2017 Oct 16;19(10):e344. [doi: 10.2196/jmir.8421]

46.   Teoh D, Shaikh R, Vogel RI, Zoellner T, Carson L, Kulasingam S, et al. A Cross-Sectional Review of Cervical Cancer Messages on Twitter During Cervical Cancer Awareness Month. Journal of Lower Genital Tract Disease 2018;22(1):8-12. [doi: 10.1097/lgt.0000000000000363]

47. Lai C, Wu W, Tsai S, Cheng S, Lin K, Liang S. The Effectiveness of a Facebook-Assisted Teaching Method on Knowledge and Attitudes About Cervical Cancer Prevention and HPV Vaccination Intention Among Female Adolescent Students in Taiwan. Health Educ Behav 2014 Dec 15;42(3):352-360. [doi: 10.1177/1090198114558591]

48. Lee KC, Oh H, Park G, Park S, Suh B, Bae WK, et al. Transmissibility of the Campaign for Colorectal Cancer Awareness in Korea Among Twitter Users. Ann Coloproctol 2016 Oct;32(5):184-189 [FREE Full text] [doi: 10.3393/ac.2016.32.5.184] [Medline: 613299883]

49. #MyBest10Seconds ECCAM2019 DiCE Campaign Report. Digestive Cancers Europe. Europa Colon;. 2019 Apr. URL: https://issuu.com/diceeurope/docs/_mybest10seconds_campaign_report_5dfcbfe1b2922b

50. Fernández-Gómez E, Díaz-Campo J. Comunicación sobre el cáncer en Facebook. Las asociaciones de Argentina, Chile, Colombia y España. cuadernos.info 2016 May 05(38):35-50. [doi: 10.7764/cdi.38.926]

51. Vraga EK, Stefanidis A, Lamprianidis G, Croitoru A, Crooks AT, Delamater PL, et al. Cancer and Social Media: A Comparison of Traffic about Breast Cancer, Prostate Cancer, and Other Reproductive Cancers on Twitter and Instagram. Journal of Health Communication 2018 Jan 09;23(2):181-189. [doi: 10.1080/10810730.2017.1421730]

52. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging Social Media to Promote Public Health Knowledge: Example of Cancer Awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17 [FREE Full text] [doi: 10.2196/publichealth.5205] [Medline: 27227152]

53. Theiss SK, Burke RM, Cory JL, Fairley TL. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. mHealth 2016 Nov 07;2:41-41 [FREE Full text] [doi: 10.21037/mhealth.2016.10.02] [Medline: 27840816]

54. Alanzi TM, Alobrah A, Alhumaidi R, Aloraifi S. Evaluation of the SnapChat mobile social networking application for breast cancer awareness among Saudi students in the Dammam Region of the Kingdom of Saudi Arabia. BCTT 2018 Jul;Volume 10:113-119. [doi: 10.2147/bctt.s166135]

55. Abramson K, Keefe B, Chou WS. Communicating about cancer through Facebook: a qualitative analysis of a breast cancer awareness page. J Health Commun 2015 Dec;20(2):237-243. [doi: 10.1080/10810730.2014.927034] [Medline: 25495200]

56. Bravo CA, Hoffman-Goetz L. Tweeting About Prostate and Testicular Cancers: Do Twitter Conversations and the 2013 Movember Canada Campaign Objectives Align? J Cancer Educ 2016 Jun;31(2):236-243. [doi: 10.1007/s13187-015-0796-1] [Medline: 25649663]

57. Bravo CA, Hoffman-Goetz L. Tweeting About Prostate and Testicular Cancers: What Are Individuals Saying in Their Discussions About the 2013 Movember Canada Campaign? J Cancer Educ 2016 Sep;31(3):559-566. [doi: 10.1007/s13187-015-0838-8] [Medline: 25903054]

58. Bravo CA, Hoffman-Goetz L. Social Media and Men's Health: A Content Analysis of Twitter Conversations During the 2013 Movember Campaigns in the United States, Canada, and the United Kingdom. Am J Mens Health 2017 Nov;11(6):1627-1641. [doi: 10.1177/1557988315617826] [Medline: 26669771]

59. Chung JE. Retweeting in health promotion: Analysis of tweets about Breast Cancer Awareness Month. Computers in Human Behavior 2017 Sep;74:112-119. [doi: 10.1016/j.chb.2017.04.025]

60. Diddi P, Lundy LK. Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. Journal of Health Communication 2017 Feb 19;22(3):243-253. [doi: 10.1080/10810730.2016.1266716]

61. Using Facebook to improve attendance at screening appointments. Good Things Foundation. 2018. URL: https://drive.google.com/file/d/1pASaOsKqsJ5AEdwT3-xAzUbZFlDI4W7K/view

62. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]

63. Adewuyi E, Adefemi K. Behavior Change Communication Using Social Media: A Review. International Journal of Communication and Health 2016;9:109-116 [FREE Full text]

64. Althoff T, Jindal P, Leskovec J. Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior. 2017 Feb Presented at: Tenth ACM International Conference on Web Search and Data Mining; February, 2017; Cambridge United Kingdom p. 537-546 URL: http://europepmc.org/abstract/MED/28345078 [doi: 10.1145/3018661.3018672]

65. Briscoe C, Aboud F. Behaviour change communication targeting four health behaviours in developing countries: a review of change techniques. Soc Sci Med 2012 Aug;75(4):612-621. [doi: 10.1016/j.socscimed.2012.03.016] [Medline: 22541798]

66. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]

67. Redaniel M, Martin R, Ridd M, Wade J, Jeffreys M. Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: historical cohort study using the Clinical Practice Research Datalink. PLoS One 2015;10(5):e0126608 [FREE Full text] [doi: 10.1371/journal.pone.0126608] [Medline: 25933397]

68. Marlow LAV, Waller J, Wardle J. Does lung cancer attract greater stigma than other cancer types? Lung Cancer 2015 Apr;88(1):104-107. [doi: 10.1016/j.lungcan.2015.01.024] [Medline: 25704958]

XSL•FO
RenderX

69. Carter-Harris L, Hermann CP, Schreiber J, Weaver MT, Rawl SM. Lung cancer stigma predicts timing of medical help-seeking behavior. Oncol Nurs Forum 2014 May;41(3):E203-E210 [FREE Full text] [doi: 10.1188/14.ONF.E203-E210] [Medline: 24769603]

70. Metwally O, Blumberg S, Ladabaum U, Sinha SR. Using Social Media to Characterize Public Sentiment Toward Medical Interventions Commonly Used for Cancer Screening: An Observational Study. J Med Internet Res 2017 Jun 07;19(6):e200 [FREE Full text] [doi: 10.2196/jmir.7485] [Medline: 28592395]

71. Vrinten C, Gallagher A, Waller J, Marlow LAV. Cancer stigma and cancer screening attendance: a population based survey in England. BMC Cancer 2019 Jun 11;19(1):566 [FREE Full text] [doi: 10.1186/s12885-019-5787-x] [Medline: 31185949]

72. Hirst Y, Stoffel S, Baio G, McGregor L, von Wagner C. Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. Eur J Cancer 2018 Nov;103:267-273 [FREE Full text] [doi: 10.1016/j.ejca.2018.07.135] [Medline: 30196989]

73. Singh G, Jemal A. Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950-2014: Over Six Decades of Changing Patterns and Widening Inequalities. J Environ Public Health 2017;2017:2819372 [FREE Full text] [doi: 10.1155/2017/2819372] [Medline: 28408935]

74. Viswanath K, Emmons K. Message Effects and Social Determinants of Health: Its Application to Cancer Disparities. Journal of Communication 2006 Aug 04;56(suppl_1):S238-SS64 [FREE Full text] [doi: 10.1111/j.1460-2466.2006.00292.x]

75. Bandura A. Social Cognitive Theory of Mass Communication. Media Psychology 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]

76. Adults: Media Use and Attitudes Report 2019. Ofcom. 2019 May 30. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf

77. Tabaac AR, Sutter ME, Wall CSJ, Baker KE. Gender Identity Disparities in Cancer Screening Behaviors. Am J Prev Med 2018 Mar;54(3):385-393. [doi: 10.1016/j.amepre.2017.11.009] [Medline: 29338956]

78. Connolly D, Hughes X, Berner A. Barriers and facilitators to cervical cancer screening among transgender men and non-binary people with a cervix: A systematic narrative review. Prev Med 2020 Jun;135:106071. [doi: 10.1016/j.ypmed.2020.106071] [Medline: 32243938]

Abbreviations

BCAM: breast cancer awareness month
BCT: behavior change technique
CINAHL: Cumulative Index to Nursing and Allied Health Literature
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

Review

Impact of Remote Consultations on Antibiotic Prescribing in Primary Health Care: Systematic Review

Seung Min Han[1], BSc; Geva Greenfield[2], PhD; Azeem Majeed[2], MB, Bch, MD, FRCP, FRCGP, FFPH; Benedict Hayhoe[2], BSc, MBBS, LLM, MD, DRCOG, DPMSA, MRCGP, FHEA, FFCI

[1]School of Public Health, Imperial College London, London, United Kingdom

[2]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Geva Greenfield, PhD
Department of Primary Care and Public Health
School of Public Health
Imperial College London
314 The Reynolds Building
St Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 02075948595
Email: g.greenfield@ic.ac.uk

Abstract

Background: There has been growing international interest in performing remote consultations in primary care, particularly amidst the current COVID-19 pandemic. Despite this, the evidence surrounding the safety of remote consultations is inconclusive. The appropriateness of antibiotic prescribing in remote consultations is an important aspect of patient safety that needs to be addressed.

Objective: This study aimed to summarize evidence on the impact of remote consultation in primary care with regard to antibiotic prescribing.

Methods: Searches were conducted in MEDLINE, Embase, HMIC, PsycINFO, and CINAHL for literature published since the databases' inception to February 2020. Peer-reviewed studies conducted in primary health care settings were included. All remote consultation types were considered, and studies were required to report any quantitative measure of antibiotic prescribing to be included in this systematic review. Studies were excluded if there were no comparison groups (face-to-face consultations).

Results: In total, 12 studies were identified. Of these, 4 studies reported higher antibiotic-prescribing rates, 5 studies reported lower antibiotic-prescribing rates, and 3 studies reported similar antibiotic-prescribing rates in remote consultations compared with face-to-face consultations. Guideline-concordant prescribing was not significantly different between remote and face-to-face consultations for patients with sinusitis, but conflicting results were found for patients with acute respiratory infections. Mixed evidence was found for follow-up visit rates after remote and face-to-face consultations.

Conclusions: There is insufficient evidence to confidently conclude that remote consulting has a significant impact on antibiotic prescribing in primary care. However, studies indicating higher prescribing rates in remote consultations than in face-to-face consultations are a concern. Further, well-conducted studies are needed to inform safe and appropriate implementation of remote consulting to ensure that there is no unintended impact on antimicrobial resistance.

(J Med Internet Res 2020;22(11):e23482)   doi:10.2196/23482

Introduction

Recent years have seen unsustainable workload increases in primary health care. Remote consultations, in which primary care professionals (PCPs) communicate with patients by telephone or internet as an alternative to face-to-face consultations, have been implemented to maximize the efficiency of primary care services and meet patient demand

for greater and more convenient access to primary health care advice [1]. Over 90 countries have been reported to be already delivering health care services over the telephone in 2016 [1], and remote consultations have been playing a substantial role in the health care response to the current COVID-19 pandemic by supporting continued access to services with minimized risk of disease transmission [2,3].

The increasingly commonplace nature of remote consulting in primary care notwithstanding, there remains much uncertainty regarding the safety and effectiveness of remote consultations [4,5]. For example, PCPs are more likely to prescribe medications in remote consultations than in face-to-face settings [6,7]. In particular, antibiotic prescribing behavior can be influenced by nonclinical factors that are unique to remote consultations, such as the inability to physically examine the patient [8]. The overprescribing of antibiotics drives antimicrobial resistance, which is a global concern with consequent impact on patients and health systems, especially primary health care. Over 80% of all antibiotic prescriptions are dispensed by primary care in the United Kingdom [9-11].

Evidence for the impact of remote consultations on antibiotic prescribing in primary care is currently unclear. Given the growing international adoption of remote consultations into primary care, which has intensified during the ongoing COVID-19 pandemic, any impact of remote consultations on antibiotic prescribing needs to be properly understood. This study aims to summarize the impact of remote consultations on primary care antibiotic prescribing.

Methods

This systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [12].

Eligibility Criteria

Included studies were required to be conducted in primary health care settings only. Any type of remote consultations between patient and PCP were accepted as the intervention. For comparison, any form of face-to-face consultation in primary health care was accepted. The primary outcomes of interest were quantitative measures of antibiotic prescriptions in remote consultations. The proportion of guideline-concordant antibiotic prescriptions and follow-up visit rates were considered secondary outcomes. Studies assessing the prescribing rates of any drug without providing data on antibiotics were excluded.

We included peer-reviewed primary research articles written in English in this systematic review. Studies could be observational or randomized controlled trials (RCTs), but conference abstracts, editorials, and qualitative studies that did not provide measures of antibiotic prescribing frequency were excluded. We did not exclude literature based on publication date so that we could capture all available literature.

Search Strategy

A scoping review was conducted to establish the search terms. A research librarian was consulted for guidance regarding the

search strategy. An initial list of search terms was developed and applied to MEDLINE and Embase to check the relevance of results, and reference lists from several relevant studies and similar reviews were manually searched to expand the search terms. Search strings were then amended according to the subject headings for each database. The final list of search terms for each database is presented in Multimedia Appendix 2.

Study Selection

A search was conducted on February 14, 2020, in 5 electronic databases: MEDLINE, Embase, PsycINFO, HMIC, and CINAHL. We had 3 reviewers independently screen studies for inclusion. One reviewer (SH) screened all titles and abstracts, and BH and GG each screened 50% of the titles and abstracts. The same approach was subsequently performed for full-text screening. Discrepancies were resolved through discussion between the reviewers. The studies were stored using Mendeley reference management software, and duplicates were removed through Mendeley's deduplication function and manual searching.

Data Extraction

Study characteristics and outcomes were extracted using a Microsoft Excel spreadsheet. The data extraction form was created in advance and finalized after piloting it on two studies. Data from included studies were extracted by SH.

Risk of Bias

The National Heart, Lung, and Blood Institute tool [13] was used to assess observational studies, and the Cochrane risk of bias tool [14] was applied to RCTs. The reviewers carefully considered the efforts required to minimize the risk of bias for each domain and ensured that the overall quality rating of each paper was not purely based on the tally of each appraisal form. SH assessed the quality of all papers and BH and GG each assessed half of the studies independently. Conflicting assessments and overall risk of bias were determined through discussion.

Analysis

Results were presented as reported by the original authors of each study, and similar outcomes that were reported frequently were grouped together for analysis. The results were presented through narrative synthesis [15]. The included studies were considered too heterogeneous in terms of study population, type of consultation, and outcome definitions for meta-analysis.

Results

Study Selection

Our electronic database search yielded 2427 results, of which 12 studies were included in this review (Figure 1) [12]. We found 2 papers [16,17] that were part of the same study with identical study periods and participants, so the results most relevant to this systematic review were reported.

Figure 1. Literature search flow chart adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart.
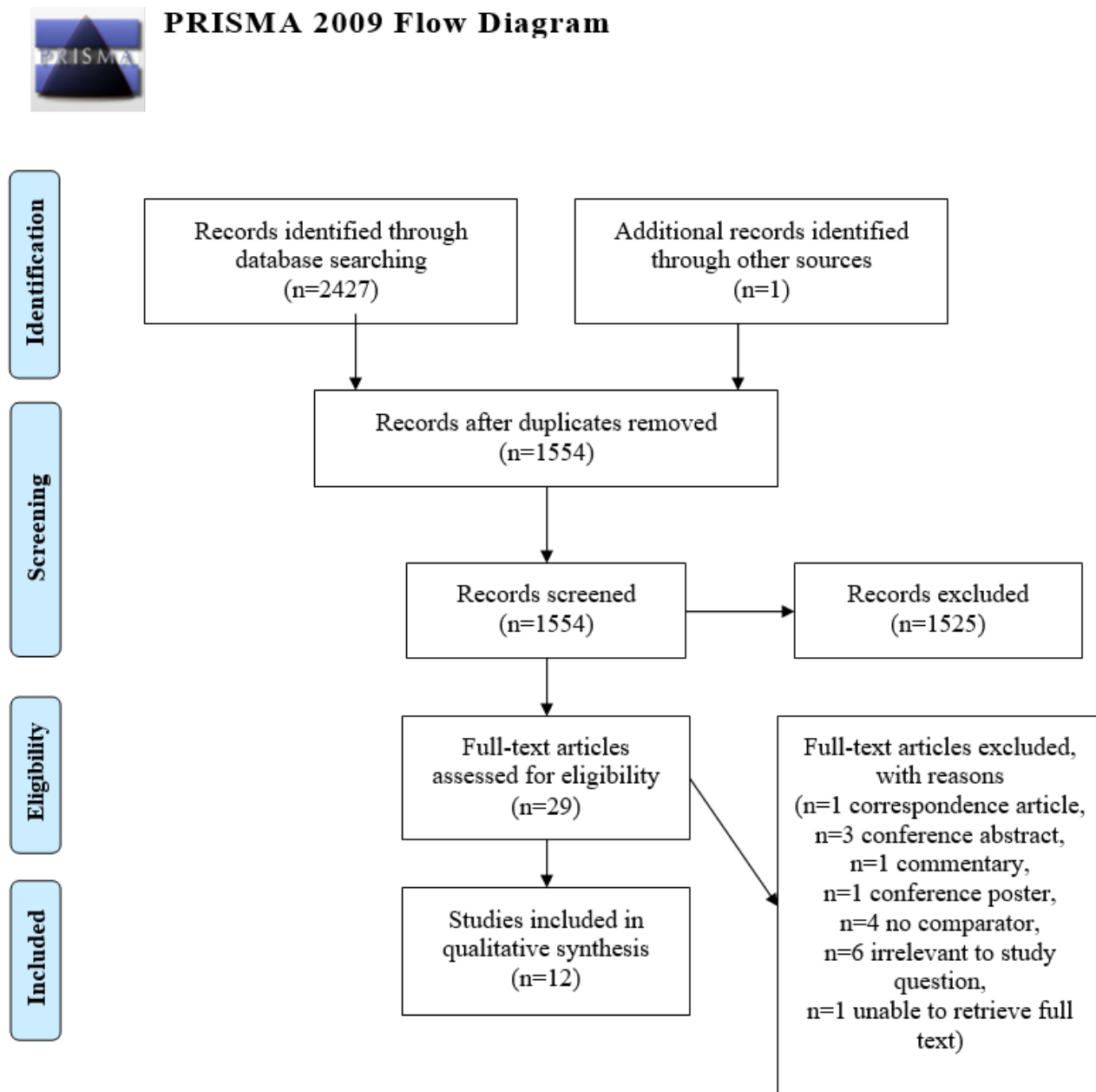
Characteristics of Included Studies

The characteristics and results of the 12 studies are summarized in Tables 1 and 2 [16-27]. Of these 12 studies, 9 were conducted in the United States [18,19,22-27], 2 were conducted in Denmark [16,17], and 1 was conducted in Norway [20]. Most studies (11/12) [16-20,22-27] had a cohort design, including 1 RCT [21].

Table 1. Summary of study characteristics.

Study (author, year)	Country	Setting	Study design	Population	Remote consultation type
Rokstad and Straand, 1997 [20]	Norway	Primary care practices	Prospective cohort	All registered patients	Telephone, letter, or through a messenger (mixed)
McKinstry et al, 2002 [21]	Scotland	Primary care practices	Randomized controlled trial	All patients that phoned for a same-day appointment	Telephone
Mehrotra et al, 2012 [22]	United States	Primary care practices	Retrospective cohort	All patients	Text-based e-visit
Huibers et al[a], 2014 [16]	Denmark	Out-of-hours primary care contacts	Retrospective cohort	All registered patients	Telephone
Ewen et al, 2015 [23]	United States	Primary care practices	Retrospective cohort	All registered patients	Telephone
Uscher-Pines et al, 2016 [24]	United States[b]	Primary care practices	Retrospective cohort	Adults aged 18-64 years	Telephone, video, internet, or mobile app consultation (mixed)
Christensen et al[a], 2016 [17]	Denmark	Out-of-hours primary care contacts	Retrospective cohort	All registered patients	Telephone
Shi et al, 2018 [25]	United States[b]	Primary care practices	Retrospective cohort	Adults aged 18-64 years	Audio and audio-visual conferencing (mixed)
Ray et al, 2019 [27]	United States[b]	Primary care practices	Retrospective cohort	Children aged 0-17 years	Audio-only or audio-video conferencing (mixed)
Murray et al, 2019 [18]	United States	Primary care practices or retail clinics	Retrospective cohort	Women aged 18-65 years	Telephone and text-based e-visit
Johnson et al, 2019 [26]	United States	Primary care practices	Retrospective cohort	Adults aged ≥18 years	Text-based e-visit
Penza et al, 2020 [19]	United States	Primary care retail clinics	Retrospective cohort	Children aged 18 months-18 years	Telephone and text-based e-visit

[a]Articles published on the same study.

[b]The data were sourced from national health insurance companies. Therefore, no specific setting was recorded.

Table 2. Main findings of studies.

Indication for antibiotics, study (author, year)	Remote consultations, N	Face-to-face consultations (control), N	Outcome measures	Results
Antibiotic-prescribing rate higher in remote consultations				
Mehrotra et al, 2012 [22]				
Sinusitis	475	4690	Antibiotic-prescribing rate[a]	Intervention: 99%, control: 94% ($P<.001$)
Urinary tract infection	99	2855	Antibiotic-prescribing rate[a]	Intervention: 99%, control: 49% ($P<.001$)
Uscher-Pines et al, 2016 [24]				
Uncomplicated acute bronchitis	168	7342	Antibiotic-avoidance rate[b]	Intervention: 16.7%, control: 27.9% ($P<.01$)
Ray et al, 2019 [27]				
Acute respiratory infection	4604	38408	Antibiotic-prescribing rate[a]	Intervention: 52%, control: 31% ($P<.001$)
Penza et al[c], 2020 [19]				
Conjunctivitis	101	202	Antibiotic-prescribing rate[a] during telephone consultations	Intervention: 41.6%, control: 19.8% ($P<.0001$)
Antibiotic-prescribing rate lower in remote consultations				
Rokstad and Straand, 1997 [20]				
Not specified	24983	42202	Proportion of prescriptions and antibiotics prescriptions made through each consultation type	Antibiotic-prescribing rate[a] lower in intervention* Intervention: 43.5% of all prescriptions; 7.8% of remote prescriptions were antibiotics Control: 56.6% of all prescriptions; 17.8% of face-to-face prescriptions were antibiotics
Huibers et al[d], 2014 [16]				
Not specified	382748	180032	Antibiotic-prescribing rate[a]	Intervention: 26.1% (95% CI 25.9-26.3) Control: 10.7% (95% CI 10.6-10.8)
Ewen et al, 2015 [23]				
Not specified	61707[e]	61707[e]	Proportion of antibiotics prescriptions out of all prescriptions	12.4% of all antibiotics prescriptions made through telephone (6617 telephone consultations and 27,487 office consultations; 63,418 antibiotics were prescribed during 61,707 consultations to 31,302 individuals)*
Shi et al, 2018 [25]				
Acute respiratory infection	38839	942163	Antibiotic-prescribing rate[a]	Intervention: 52%, control: 53% ($P<.01$)
Johnson et al, 2019 [26]				
Sinusitis	175	175	Antibiotic-prescribing rate[a]	Intervention: 68.6%, control: 94.3% ($P<.001$)
No significant difference in antibiotic-prescribing rate				
McKinstry et al, 2002 [21]				

Indication for antibiotics, study (author, year)	Remote consultations, N	Face-to-face consultations (control), N	Outcome measures	Results
Not specified	187	181	Antibiotic-prescribing rate[a]	Intervention: 19.3%, control: 16.0%; difference: −3.3% (95% CI −11.1% to 4.5%)
Murray et al, 2019 [18]				
Urinary tract infection	150	150	Antibiotic-prescribing rate[a] in telephone consultations	Intervention: 81%, control: 83% (P=.76)
Urinary tract infection	150	150	Antibiotic-prescribing rate[a] from text-based e-visits	Intervention: 81%, control: 83% (P=.65)
Penza et al[c], 2020 [19]				
Conjunctivitis	101	202	Antibiotic-prescribing rate[a] from text-based e-visits	Intervention: 25.7%, control: 19.8% (P=.24)

[a]Antibiotic-prescribing rate: percentage of consultations that resulted in antibiotic management per consultation type.

[b]Antibiotic-avoidance rate: percentage of patients that did not receive antibiotics for uncomplicated acute bronchitis, as they had no clinical indication.

[c]Results of the same study for different populations reported separately.

[d][16] and [17] are articles published on the same study. The results from the Huibers et al [16] study are reported in this table.

[e]Number of remote consultations and face-to-face consultations in this study was not available. The number of consultations altogether has been reported instead.

*P values or confidence intervals not reported in the original studies.

The study population varied in age and sex depending on whether the study investigated specific conditions. Of 2 studies analyzing antibiotic prescribing for urinary tract infection (UTI) [18,22], 1 limited their participants to adult women [18]. Additionally, 2 studies focused only on children with conjunctivitis or acute respiratory infection [19,27].

A majority of studies (7/12) employed telephone consultations, text-based e-visits, or both as their intervention [17-19,21-23,26]. Moreover, 2 studies [25,27] grouped audio-only and video consultations for their intervention, and another study [20] grouped all consultations made through telephone, letters, or messengers. The control arm was face-to-face primary care clinic consultations for all except 2 studies. The Penza et al [19] control group was made up of walk-in retail clinic patients, and the controls for Murray et al [18] were retail clinic and primary care practice patients. Furthermore, while most consultations were evaluated by primary care physicians, telephone consultations in the Penza et al [19] and Murray et al [18] studies were evaluated by registered nurses. Text-based e-visits were assessed by advanced practice providers in the Murray et al [18] study and nurse practitioners in the Penza et al [19] study.

Risk of Bias Within Studies

The quality assessment is presented in Multimedia Appendix 3. The quality of studies was generally fair, with 7 moderate-quality studies, including the RCT, and 5 observational studies that were considered high quality. All studies had clearly defined their objectives, study population, and exposure and outcome measures. However, only 2 studies provided sample size justifications or power calculations [17,26]. Additionally, 4 studies lacked adjustment for any confounding factors [16,18,20,23]. The RCT described adequate randomization, but lacked blinding [21]. However, we did not consider this to significantly impact the quality, since blinding to consultation type is rarely feasible.

Impact of Remote Consultations on Antibiotic Prescribing Behavior

The main findings are outlined in Table 2, with confidence intervals and P values reported where available. In most studies, the impact of remote consultation on antibiotic prescribing was measured by the percentage of consultations with one or more antibiotic prescriptions for each type of consultation. This measure has been uniformly referred to as the antibiotic-prescribing rate in this review, following the practice in the included studies.

We found 3 studies that reported the antibiotic-prescribing rate to be higher in remote settings than face-to-face consultations [19,22,27]. The difference in prescribing rate ranged from 5% (n=201) to 50% (n=5165) [22]. Penza et al found this relationship for both telephone and text-based e-visits, but only the difference in telephone consultations was statistically significant [19]. Another study concluded that fewer antibiotics were avoided for bronchitis patients consulting over the phone; patients were more likely to receive an antibiotic for the same condition when consulting remotely [24].

We also found 3 cohort studies that reported patients were more likely to be prescribed an antibiotic through face-to-face consultations [16,25,26]. The differences in prescribing rates were generally smaller in these studies; 1 study found a 1% difference (n=1,336,867) [25], and the largest difference observed was 25.7% (n=350) [26]. Additionally, 2 moderate-quality studies, an RCT and a retrospective cohort

study, noted no significant difference in prescribing rate between the different settings [18,21].

It was difficult to compare 2 studies with the other included studies due to differences in outcome measures; neither included a direct estimate of antibiotic prescribing. Rokstad and Straand [20] and Ewen et al [23] claimed that antibiotic-prescribing rate was higher in face-to-face consultations based on the percentage of antibiotic prescriptions issued through each consultation type. However, their results do not consider multiple prescriptions made in one consultation. Therefore, their claims should be considered with caution.

Clinicians may be more likely to prescribe antibiotics for UTI in remote consultations. Of 2 studies assessing patients with suspected or confirmed UTI, 1 study found higher antibiotic-prescribing rates in remote consultations [22]. Mehrotra et al [22] concluded that PCPs were less likely to order a UTI-relevant test in face-to-face consultations (did not order: 8%, did order: 51%; P<.01), but more likely to prescribe an antibiotic for UTI in remote consultations. Another moderate-quality study found no significant difference in antibiotic-prescribing rate for patients with UTI [18]. However, the researchers hypothesized that their results could have been affected by differences in face-to-face settings; face-to-face consultations in the Mehrotra et al [22] study were performed in retail clinics where there may have been less continuity of care than primary care practices.

Results for respiratory infections were mixed. In 3 studies, of which 2 were of high quality, researchers found higher antibiotic-prescribing rates in remote consultations than in direct PCP consultations [22,24,27]. However, 2 other high-quality studies noted lower antibiotic-prescribing rates for patients consulting remotely [25,26].

Impact of Remote Consultations on Guideline-Concordant Prescribing Rate

We found 4 observational studies that reported guideline-concordant prescribing rates or guideline-recommended prescribing rates, which measured the appropriateness of the prescriptions against local or national guidelines [22,25-27] (Table 2). All studies were from the United States and investigated populations with a confirmed diagnosis of sinusitis [22,26], UTI [22], or acute respiratory infection [25,27].

In contrast to the findings for antibiotic-prescribing rate, the guideline-concordant antibiotic management for sinusitis and patients with UTI revealed no significant difference between remote and face-to-face consultations [22,26]. However, conflicting results were observed for patients with acute respiratory infection [25,27].

Impact of Remote Consultations on Follow-Up Visits

We found 5 US-based retrospective cohort studies that investigated follow-up visit rates after initial consultation for the same presentation [18,19,22,25,26]. Of these 5 studies, 3 found that patients who were seen remotely were more likely to have another follow-up visit than those who attended face-to-face consultations [19,25,26]. Shi et al [25] found this

to be true for those who followed up within 2-21 days after their first visit. Results from the Penza et al [19] study indicated higher follow-up rates in both e-visits and telephone consultations than in face-to-face consultations 7 days after the initial visit. Johnson et al [26] found this relationship to be true for text-based e-visits relating to sinusitis in the subsequent 24 hours and 30 days after the initial consultation, but found no difference in follow-up consultations at 7 days after the initial consultation. However, Murray et al [18] and Mehrotra et al [22] found no significant difference in follow-up visit rates between the consultation types within the following 3 weeks and 30 days after the initial consultation, respectively. Further, Murray et al [18] found no significant difference in antibiotic-prescribing rates in the initial consultation for patients who were followed up with.

Discussion

Principal Results

To our knowledge, this is the first systematic review to examine how antibiotic prescribing is affected by remote consultation in primary care. This review of moderate- to high-quality studies found that evidence regarding the impact of remote consultations on antibiotic prescribing was mixed. Studies reporting higher antibiotic-prescribing rates in remote consultations than in face-to-face consultations were generally of better quality. However, the inconsistency of results and the small number of studies make it difficult to draw strong conclusions for the effect of remote consultations on antibiotic prescribing.

The studies examining specific indications for antibiotics suggested that antibiotic-prescribing rates for patients with UTI in remote consultations was higher than in face-to-face consultations, but 1 study did not find any difference [18]. Guideline-concordant prescribing rates for patients with UTI or sinusitis were not significantly changed by remote consultations. However, there was mixed evidence regarding whether remote consultations were more likely to be followed up with another consultation for the same condition.

Limitations

We tried to conduct a comprehensive search by manually searching reference lists to find relevant search terms for remote consultation. However, there is significant variation in terminology among researchers, and this could have led to the omission of a few relevant papers. Other challenges we faced while conducting this review included the dearth of relevant papers. The included studies were conducted in 1 of 4 high-income countries. Given the growing attention on remote consultations in low- and middle-income countries [1], it is possible that relevant papers in grey literature or papers written in languages other than English exist, but were excluded due to our selection criteria. Additionally, the effect of the setting of studies is unaccounted for. For example, retail clinics differ from primary care practices, as there is less chance of establishing a long-term doctor-patient relationship. Moreover, in some studies, remote consultations were evaluated by nurse practitioners, who may have had a different skillset compared to primary care physicians who consulted with patients

XSL•FO

RenderX

face-to-face. With the small number of studies, no clear pattern emerged in terms of the impact of setting on the outcome.

Comparison With Prior Work

Compared to face-to-face consultations, PCPs order fewer tests and investigations and are unable to physically examine patients in remote settings, which can affect the appropriateness of patient management [8,28]. Clinicians consulting in retail clinics via private telemedicine providers could feel pressured to prescribe antibiotics due to the expectations of patients who pay to see clinicians and the diagnostic uncertainty that stems from the lack of continuity in these commercial remote consultations [29]. On the other hand, Banks et al [30] noted that many remote consultations were followed up with face-to-face appointments for adequate clinical assessment. As antibiotic prescriptions made in follow-up appointments contribute to prescriptions made in face-to-face consultations, this could provide a partial explanation for why some of the included studies found lower antibiotic-prescribing rates in initial remote consultations and higher frequencies of follow-up consultations [25,26].

However, research on the remote prescribing of antibiotics is too limited to make useful comparisons. A systematic review of reviews on the benefits of telemedicine in 2002 found too little high-quality evidence to confidently conclude that telemedicine was beneficial. However, the review reported that research was beginning to address the literature gaps in the field [31]. More recent reviews agree that there is a need for more research on the safety of care [4] and the quality and safety of prescribing through remote consultations [4,32]. As this is a novel study, it is difficult to compare the findings of this review directly with other reviews. We hope that this review serves as a reference point for future studies.

Implication for Research and Practice

We anticipate that remote consultations will continue to be used frequently following the COVID-19 pandemic, leading to a different mix of cases involving remote consultations compared to the mix of cases seen prior to the pandemic. Future studies should focus on conducting trials that adjust for this difference in case makeup. Moreover, the current literature is reliant on studies from high-income countries and observational studies, as confirmed by the inclusion of only one RCT in this review. Randomized trials from a variety of geographical settings are needed to achieve balance in the literature.

Antibiotic prescribing should only occur when it is safe, clinically indicated, and likely to be of benefit, regardless of consultation type [33]. Consequently, the quality and safety of antibiotic prescribing in remote consultations should be comparable to that of face-to-face consultations. Similar guideline-concordant prescribing between remote and face-to-face consultations is reassuring, but the evidence for whether the management at remote consultations is effective is less clear. Follow-up visits after remote consultations are often necessary per clinicians' advice and patients' need for physical examinations. Therefore, a high follow-up visit rate does not necessarily correlate with poor management at initial consultation, but it does raise the question of the effectiveness (including cost-effectiveness) of remote consultations. Future research investigating the resolution of symptoms as an endpoint could be beneficial.

The divided weight of evidence in this review makes it difficult to inform health policy, as the evidence regarding remote consultations is still evolving. Nonetheless, good quality evidence suggesting higher antibiotic-prescribing rate in remote consultations should not be ignored. As remote consultations are being used more frequently due to the pandemic, clear guidelines and criteria for face-to-face consultations and antibiotic management are needed. Meanwhile, PCPs should continue to be cautious when prescribing antibiotics and remain attentive to local and national guidelines. Furthermore, current antibiotic stewardship programs [34,35] could be adapted and implemented into remote care to encourage appropriate prescribing.

Conclusions

We found inconsistent evidence across the included studies for the impact of remote consulting on antibiotic prescribing in primary care. However, as the use of remote consultations continues to increase in primary care worldwide, ensuring the safety and quality of these consultations, including avoiding adverse impacts on antimicrobial resistance, should be prioritized. Studies indicating higher prescribing rates than face-to-face consulting are a concern, and PCPs should be cautious when considering prescribing antibiotics through remote consultations. Randomized trials are needed in a variety of geographical settings to inform policy on the wide-scale implementation of remote consultations. This type of research should be a priority, as long-term increases in remote consulting in primary care seems to be an inevitable consequence of the global COVID-19 pandemic.

Disclaimer

Authors' Contributions

SH helped with refining the research question; developed the search strings; conducted screening, critical appraisal, data extraction, and data analysis; and wrote the first draft of the manuscript. BH and GG helped with the conception and design of the review and conducted screening and critical appraisal. All authors contributed to all versions of the manuscript and approved the final submitted version.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist of items to include when reporting a systematic review or meta-analysis.

[[PDF File (Adobe PDF File), 52 KB](#) - [jmir_v22i11e23482_app1.pdf](#)]

Multimedia Appendix 2

Search terms and strategy for each electronic database.

[[PDF File (Adobe PDF File), 98 KB](#) - [jmir_v22i11e23482_app2.pdf](#)]

Multimedia Appendix 3

Quality assessment for observational studies and randomized trials.

[[PDF File (Adobe PDF File), 70 KB](#) - [jmir_v22i11e23482_app3.pdf](#)]

References

1. World Health Organization. In: World Health Organization, editor. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. Geneva, Switzerland: WHO Document Production Services; Dec 15, 2016.
2. Majeed A, Maile EJ, Bindman AB. The primary care response to COVID-19 in England's National Health Service. J R Soc Med 2020 Jun;113(6):208-210. [doi: 10.1177/0141076820931452] [Medline: 32521196]
3. Alwashmi MF. The Use of Digital Health in the Detection and Management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2906 [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]
4. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic Consultation in Primary Care Between Providers and Patients: Systematic Review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
5. Armfield NR, Gray LC, Smith AC. Clinical use of Skype: a review of the evidence base. J Telemed Telecare 2012 Apr;18(3):125-127. [doi: 10.1258/jtt.2012.SFT101] [Medline: 22362829]
6. Haaijer-Ruskamp FM, Stewart R, Wesseling H. Does indirect consultation lead to overprescribing in general practice? Soc Sci Med 1987;25(1):43-46. [doi: 10.1016/0277-9536(87)90205-x] [Medline: 3616696]
7. Sundseth AC, Gjelstad S, Straand J, Rosvold EO. General practitioners' prescriptions of benzodiazepines, Z-hypnotics and opioid analgesics for elderly patients during direct and indirect contacts. A cross-sectional, observational study. Scand J Prim Health Care 2018 Jun;36(2):115-122 [FREE Full text] [doi: 10.1080/02813432.2018.1459164] [Medline: 29656692]
8. Hayhoe B, Greenfield G, Majeed A. Is it getting easier to obtain antibiotics in the UK? Br J Gen Pract 2019 Feb;69(679):54-55. [doi: 10.3399/bjgp19X700829] [Medline: 30704991]
9. Hay AD. Antibiotic prescribing in primary care. BMJ 2019 Feb 27;364:l780. [doi: 10.1136/bmj.l780] [Medline: 30814116]
10. Llor C, Bjerrum L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Ther Adv Drug Saf 2014 Dec;5(6):229-241 [FREE Full text] [doi: 10.1177/2042098614554919] [Medline: 25436105]
11. Interagency Coordination Group on Antimicrobial Resistance. No Time To Wait: Securing the Future From Drug-Resistant Infections. Report to the Secretar-General of the United Nations. World Health Organization. 2019 Apr. URL: https://www.who.int/docs/default-source/documents/no-time-to-wait-securing-the-future-from-drug-resistant-infections-en.pdf?sfvrsn=5b424d7_6 [accessed 2020-10-21]
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
13. National Heart, Lung, and Blood Institute. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Study Quality Assessment Tools. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2020-10-21]
14. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
15. Popay J, Arai L, Britten N, Petticrew M, Rodgers M, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme. ResearchGate.: Lancaster University URL: https://tinyurl.com/y6regpvh [accessed 2020-10-21]
16. Huibers L, Moth G, Christensen MB, Vedsted P. Antibiotic prescribing patterns in out-of-hours primary care: a population-based descriptive study. Scand J Prim Health Care 2014 Dec;32(4):200-207 [FREE Full text] [doi: 10.3109/02813432.2014.972067] [Medline: 25350313]

XSL•FO

RenderX

17. Christensen MB, Nørøxe KB, Moth G, Vedsted P, Huibers L. Drug prescriptions in Danish out-of-hours primary care: a 1-yearpopulation-based study. Scand J Prim Health Care 2016 Dec;34(4):453-458 [FREE Full text] [doi: 10.1080/02813432.2016.1248622] [Medline: 27804314]

18. Murray MA, Penza KS, Myers JF, Furst JW, Pecina JL. Comparison of eVisit Management of Urinary Symptoms and Urinary Tract Infections with Standard Care. Telemed J E Health 2020 May;26(5):639-644. [doi: 10.1089/tmj.2019.0044] [Medline: 31313978]

19. Penza KS, Murray MA, Myers JF, Maxson J, Furst JW, Pecina JL. Treating pediatric conjunctivitis without an exam: An evaluation of outcomes and antibiotic usage. J Telemed Telecare 2020;26(1-2):73-78. [doi: 10.1177/1357633X18793031] [Medline: 30153768]

20. Rokstad K, Straand J. Drug prescribing during direct and indirect contacts with patients in general practice: a report from the Møre & Romsdal Prescription Study. Scand J Prim Health Care 1997 Jun;15(2):103-108. [doi: 10.3109/02813439709018496] [Medline: 9232712]

21. McKinstry B, Walker J, Campbell C, Heaney D, Wyke S. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. Br J Gen Pract 2002 Apr;52(477):306-310. [Medline: 11942448]

22. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. JAMA Intern Med 2013 Jan 14;173(1):72-74 [FREE Full text] [doi: 10.1001/2013.jamainternmed.305] [Medline: 23403816]

23. Ewen E, Willey VJ, Kolm P, McGhan WF, Drees M. Antibiotic prescribing by telephone in primary care. Pharmacoepidemiol Drug Saf 2015 Feb;24(2):113-120. [doi: 10.1002/pds.3686] [Medline: 25079292]

24. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and Quality of Care in Direct-to-Consumer Telemedicine. Telemed J E Health 2016 Apr;22(4):282-287 [FREE Full text] [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]

25. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality Of Care For Acute Respiratory Infections During Direct-To-Consumer Telemedicine Visits For Adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023 [FREE Full text] [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]

26. Johnson KM, Dumkow LE, Burns KW, Yee MA, Egwuatu NE. Comparison of Diagnosis and Prescribing Practices Between Virtual Visits and Office Visits for Adults Diagnosed With Sinusitis Within a Primary Care Network. Open Forum Infect Dis 2019 Sep;6(9):ofz393 [FREE Full text] [doi: 10.1093/ofid/ofz393] [Medline: 31660415]

27. Ray KN, Shi Z, Gidengil CA, Poon SJ, Uscher-Pines L, Mehrotra A. Antibiotic Prescribing During Pediatric Direct-to-Consumer Telemedicine Visits. Pediatrics 2019 May;143(5):e20182491. [doi: 10.1542/peds.2018-2491] [Medline: 30962253]

28. Wallace P, Haines A, Harrison R, Barber J, Thompson S, Jacklin P, Virtual Outreach Project Group. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. Lancet 2002 Jun 08;359(9322):1961-1968. [doi: 10.1016/s0140-6736(02)08828-1] [Medline: 12076550]

29. DeJong C, Santa J, Dudley RA. Websites that offer care over the Internet: is there an access quality tradeoff? JAMA 2014 Apr 02;311(13):1287-1288. [doi: 10.1001/jama.2014.1026] [Medline: 24691602]

30. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8. [doi: 10.3399/bjgp17X693509] [Medline: 29109115]

31. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]

32. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]

33. Remote prescribing high level principles Internet. General Medical Council. URL: https://www.gmc-uk.org/ethical-guidance/learning-materials/remote-prescribing-high-level-principles [accessed 2020-10-21]

34. TARGET Antibiotics Toolkit Internet. Royal College of General Practitioners. URL: https://www.rcgp.org.uk/targetantibiotics [accessed 2020-10-21]

35. Choosing Wisely: Promoting conversations between providers and patients. American Board of Internal Medicine Foundation. URL: https://www.choosingwisely.org/ [accessed 2020-10-21]

Abbreviations

PCP: primary care professionals
RCT: randomized controlled trial
UTI: urinary tract infection

Review

Barriers to and Facilitators of Technology in Cardiac Rehabilitation and Self-Management: Systematic Qualitative Grounded Theory Review

Shreya Tadas[1], MSc; David Coyle[1], PhD

University College Dublin, Dublin, Ireland

Corresponding Author:
Shreya Tadas, MSc
University College Dublin
Belfield
Dublin, Dublin 4
Ireland
Phone: 353 899882380
Email: shreya.tadas@ucdconnect.ie

Abstract

Background: Dealing with cardiovascular disease is challenging, and people often struggle to follow rehabilitation and self-management programs. Several systematic reviews have explored quantitative evidence on the potential of digital interventions to support cardiac rehabilitation (CR) and self-management. However, although promising, evidence regarding the effectiveness and uptake of existing interventions is mixed. This paper takes a different but complementary approach, focusing on qualitative data related to people's experiences of technology in this space.

Objective: Through a qualitative approach, this review aims to engage more directly with people's experiences of technology that supports CR and self-management. The primary objective of this paper is to provide answers to the following research question: What are the primary barriers to and facilitators and trends of digital interventions to support CR and self-management? This question is addressed by synthesizing evidence from both medical and computer science literature. Given the strong evidence from the field of human-computer interaction that user-centered and iterative design methods increase the success of digital health interventions, we also assess the degree to which user-centered and iterative methods have been applied in previous work.

Methods: A grounded theory literature review of articles from the following major electronic databases was conducted: ACM Digital Library, PsycINFO, Scopus, and PubMed. Papers published in the last 10 years, 2009 to 2019, were considered, and a systematic search with predefined keywords was conducted. Papers were screened against predefined inclusion and exclusion criteria. Comparative and in-depth analysis of the extracted qualitative data was carried out through 3 levels of iterative coding and concept development.

Results: A total of 4282 articles were identified in the initial search. After screening, 61 articles remained, which were both qualitative and quantitative studies and met our inclusion criteria for technology use and health condition. Of the 61 articles, 16 qualitative articles were included in the final analysis. Key factors that acted as barriers and facilitators were background knowledge and in-the-moment understanding, personal responsibility and social connectedness, and the need to support engagement while avoiding overburdening people. Although some studies applied user-centered methods, only 6 involved users throughout the design process. There was limited evidence of studies applying iterative approaches.

Conclusions: The use of technology is acceptable to many people undergoing CR and self-management. Although background knowledge is an important facilitator, technology should also support greater ongoing and in-the-moment understanding. Connectedness is valuable, but to avoid becoming a barrier, technology must also respect and enable individual responsibility. Personalization and gamification can also act as facilitators of engagement, but care must be taken to avoid overburdening people. Further application of user-centered and iterative methods represents a significant opportunity in this space.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Cardiovascular diseases (CVDs) are the leading cause of death worldwide. An estimated 17.9 million people died from CVD in 2016, representing 31% of all global deaths [1,2]. By 2035, more than 130 million adults in the US population (45.1%) are projected to have some form of CVD, with the total costs of CVD expected to reach US $1.1 trillion [3]. Improved cardiovascular outcomes depend largely on how well affected people manage their condition [4]. Physical rehabilitation and lifestyle management are critical components of programs aimed at primary and secondary prevention of CVD. A major challenge in implementing these strategies is ensuring good patient engagement and compliance with prescribed exercise programs and nutrition plans. Evidence from the literature suggests that tightly supervised intervention programs are most successful and that self-directed management is less successful because of problems with engagement and adherence. The problem lies in expecting patients with a wide variety of life patterns and personality types to conform to standardized programs that do not fit with their ever-changing context [5].

After a person is hospitalized and following a discharge and recuperation period, they are typically recommended to attend a cardiac rehabilitation (CR) program offered by hospitals. Following this, they need to continue to self-manage their cardiac health. CR is considered a vital part of long-term recovery by targeting risk factor modification, supervised exercise, psychological support, and medication review [6]. However, the uptake of CR programs remains poor because of factors such as age, gender, lack of knowledge, transportation, motivation, and social support [7,8]. This also has an impact on people's subsequent ability to self-manage their condition. Barlow et al [9] state, "self-management refers to the individual's ability to manage the symptoms, treatment, physical and psychological consequences and lifestyle changes." Recent research suggests that digital health interventions can play an important role in supporting both rehabilitation and self-management. A systematic review of mobile phone apps to support self-care following heart failure by Athilingam and Jenkins [10] demonstrated positive trends and cost-effectiveness, enabling increased access to symptom monitoring and promoting patient engagement in their own homes. Similarly, a review by Piette et al [11] on mobile health (mHealth) technologies for CVD reduction and management found evidence that mHealth interventions can improve cardiovascular-related lifestyle behaviors and disease management. The authors emphasize the need for new interventions that build on evidence-based behavioral theories and are adaptive to a patient's unique and changing needs. Jörntén-Karlsson et al [12] also suggested mHealth as an effective long-term alternative to face-to-face rehabilitation and consultation, with the potential to reach more patients at a relatively lower cost. They found evidence that digital interventions can have a positive impact on patients with CVD but again stressed the need for easy to use, personalized, and user-friendly apps that can cater to patients from all age groups, especially older age groups. This recognition of the specific needs of older adults is critical, given the significant

impact of CVD among this age group. However, recognizing the potential of technology to support patients with CVD across diverse age groups is also important, given the evidence from Foster et al [13] and Andersson and Vasan [14] that CVD impacts adults in all age groups. In line with this, a survey conducted by Gallagher et al [15], to assess the use of mobile technology among different demographics, demonstrates that mobile technology, when modified to suit different subgroups, offers an important opportunity to improve access to secondary prevention for cardiac patients.

Although there is a significant literature and a growing number of reviews on digital interventions for CVD rehabilitation and management, most previous studies base their conclusions on quantitative data. To better understand what drives the effectiveness and usage of technologies, there is also a need to analyze the collective perspectives of patients, focusing on their experiences, needs, and the barriers they face in using digital interventions. The literature outlined earlier has provided evidence that personalization [16] and the application of appropriate theory will play an important role in improving digital health technologies that target CVD. For example, behavior change theories and models can help inform the design of technical systems, guide evaluation strategies, and define target users [17,18]. In addition, persuasive design patterns can be used in digital interventions to address the challenge of obtaining sustained user engagement and behavior change among patients with CVD [19]. Building on this evidence, a greater understanding of patients' experiences will provide the insight needed to design future technology and increase the success of technologies when deployed in real-world contexts. By improving adherence to lifestyle changes, appropriately designed digital health technologies that apply this insight can ultimately help to prevent recurrence of cardiac conditions.

The analysis in this paper draws strongly on research in the field of human-computer interaction (HCI). Our findings are analyzed from an HCI perspective, which emphasizes the benefits of iterative development of technology and user involvement throughout the design and evaluation process [20-23]. HCI approaches have been successfully applied to rehabilitation and self-management in other health domains [24-27]. Our decision to focus on both rehabilitation and self-management followed multiple discussions among the authors and cardiologists, which reflected the degree to which these issues are interconnected. The papers selected in this review have dealt with some of the common issues and challenges. An overview of these interventions, along with the synthesis of patients' experiences, can be beneficial to both medical and HCI researchers. To the best of our knowledge, no previous systematic review has combined qualitative review methods and an HCI perspective to identify challenges and opportunities in the design of technology to support CR and self-management.

Objectives

The primary objective of this paper is to provide answers to the following research question: What are the primary barriers to and facilitators and trends of digital interventions to support CR and self-management? This question is answered by synthesizing evidence from both medical and computer science

literature. Using a qualitative approach, we aim to engage more directly with people's needs from and experiences of technology that supports CR and self-management. Given the strong evidence from the field of HCI that user-centered and iterative design methods increase the success of digital health interventions, we also assess the degree to which user-centered and iterative methods have been applied in the studies included in this review.

This review follows the grounded theory literature review (GTLR) method [28]. GTLR aims at producing new insights and enables researchers to develop concept-centric yet accurate reviews through a 5-stage iterative process. The GTLR method adopts a rigorous search and selection process, eventually invoking the grounded theory method for the analysis stage. GTLR recommends that initial research questions are identified at the outset of the review process and allows for a bottom-up iterative approach in which new concepts are identified via a thorough and progressive analysis. Initial questions help focus on the review during the selection and analysis stages, but based on concepts identified during the analysis stage, it is acceptable for the final concepts to differ somewhat in focus from the initial questions. Following multiple rounds of discussion and refinement among the authors and cardiologists involved in this project, the following initial research questions were identified:

1. What kind of technological support is provided for CR and self-management?
2. What design approaches were applied in designing the technologies identified?
3. What experiences and attitudes do patients have of technology?
4. What are the barriers to using technology for rehabilitation and self-management after a cardiac incident?
5. What are the facilitators for using technology for rehabilitation and self-management after a cardiac incident?

Methods

Overview

This review follows the 5 stages recommended in the GTLR method [28]: (1) identifying the key research questions, appropriate sources, and search terms; (2) search for potential papers; (3) defined filtering for selection of papers and refining the sample for review; (4) a comparative and in-depth analysis of the papers through 3 coding levels; and (5) representing the emerging categories and concepts. In addition, we used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) as guidance for conducting this review. The complete PRISMA checklist for this paper is included in Multimedia Appendix 1.

In this section, we describe the inclusion and exclusion criteria of our review, database sources and search keywords used, the screening and selection process, data extraction process, and, finally, the analysis process.

Search Strategy

To include a wide range of perspectives on designing technologies for rehabilitation and self-management of cardiac conditions, we selected papers from PsycINFO, Scopus, PubMed, and ACM (Association for Computing Machinery) Digital Library. HCI literature about designing technology for cardiac conditions was gathered from the ACM Digital Library. Similarly, psychology and medical literature on these types of technologies were gathered from PsycINFO and PubMed. Other major journals and conferences, such as Biomed Central, IEEE (Institute of Electrical and Electronics Engineers), BMJ (British medical journal), International Journal of Telemedicine and Applications, SAGE (Scientific Advisory Group for Emergencies), and Global Telehealth, are included in Scopus.

Title, abstract, and keyword searches were carried out on the above mentioned databases to obtain the results for this review (Multimedia Appendix 2). On the basis of the studies we were familiar with and to follow a structured process to define the keywords, we selected keywords to address 3 areas: domain, technology, and intervention that we considered most relevant to identify papers of interest (Textbox 1). Domain keywords focused on CVD as the main field interest, together with related medical terms (eg, coronary artery). Technology keywords addressed diverse technologies used in inventions (eg, mobile phones, sensors, and telehealth). Intervention keywords reflect the different types of interventions addressing the field of CVD (eg, tracking and behavior change). It is important to note that our search strings include both Medical Subject Headings (MeSH) and non-MeSH terms. This decision was made because the study aimed at a broad exploration of research in both technology (eg, HCI and software engineering) and medical disciplines. The technology databases included in our study (eg, the ACM Digital Library) do not recognize MeSH terms. Including both MeSH and non-MeSH terms represented the most balanced approach and helped to ensure consistency of search terms across the different databases.

We limited our search to papers published in the last 10 years and focused on papers in the English language and including adult patients.

Textbox 1. Keywords used in the search terms.

Domain

- Cardiovascular disease
- Cardiology
- Cardiac
- Heart disease
- Coronary heart disease
- Coronary artery disease
- Heart failure

Technology

- Mobile
- Wearable
- Wearable sensors
- mHealth interventions
- Smartphone
- Tele-monitoring
- Sensing system
- Telehealth
- Telemedicine

Intervention

- Persuasive or persuasion
- Quantified self
- Tracking
- Behavior change or behavior
- Personal informatics
- Habit
- Prevention
- Detection
- Rehabilitation
- Management

Eligibility Criteria

The review was concerned with the use of technology for self-management and rehabilitation practices in the context of CVDs. This excluded several papers that would otherwise be featured in the review, such as those suggesting design concepts without evaluating them [29,30], those describing algorithms or software architectures to solve specific self-care problems [31,32], and those focusing on monitoring and detection techniques to support primary prevention of CVD [33,34]. These types of studies are very relevant to CVD in a broader sense, but as they do not provide evidence on the use of technology to support self-management or rehabilitation, they were excluded from the review. The papers included in this review involved an active role for patients living with cardiac conditions and technology that could be controlled by the patients rather than those in which patients have a more passive role. This meant excluding a number of technologies used only in clinical settings and technologies based on biomarkers, photoplethysmogram, implantable devices, and defibrillators. Excluding them enabled us to focus on the lived experience of people with CVD, rather than the clinical context of care.

Furthermore, this review focuses on studies of patients with cardiac conditions. This excluded self-management and rehabilitation technologies focusing on other chronic conditions [19], wellness and lifestyle [35,36], or quantifying habits for health [37,38]. By keeping the focus on cardiac conditions, the motivation for using the technology was to maintain cardiac health, not to pursue personal interest, leisure, or general well-being, which would likely bring different principles for design and use. To attain subjective perspectives of patients' needs and seek answers to our research questions, we focused

on qualitative study methods for this review. Therefore, to be eligible for inclusion in this review, papers needed to include a technology intervention for cardiac management or rehabilitation, use qualitative study methods, and describe the use and evaluation of technology with users. Papers that did not follow the criteria were rejected. The inclusion and exclusion criteria are listed in Textboxes 2 and 3, respectively.

Textbox 2. Inclusion criteria.

Domain

- Cardiac condition

Technology

- Use of technology with evaluation
- Technologies having active patient role (eg, mobile, wearable, mobile health, and telemedicine)

Intervention

- Secondary prevention involving self-management and rehabilitation

Textbox 3. Exclusion criteria.

Domain

- Other chronic conditions and general well-being and lifestyle

Technology

- Design concepts, technology description, algorithms, and software architecture without evaluation
- Technologies having passive patient role: biomarkers
- Technology used in clinical settings
- Photoplethysmogram
- Implantable devices
- Defibrillators

Intervention

- Detection and monitoring for primary prevention

Screening and Data Extraction

The search keywords retrieved 4282 articles, of which 3973 remained after removing duplicates. We first performed a prescreening of these papers by reading the title and abstract and removed papers concerning research abstracts, systematic reviews, protocols, workshops, studies dealing with patients aged <18 years, studies involving chemical and biological sciences, and studies involving clinical procedures. At this stage, the first author (ST) reviewed all papers, and the second author was consulted in any situation where the first author was uncertain. Where any disagreement occurred, the paper was not excluded at this stage. In the second phase of screening, the first author reviewed the title and abstract of all remaining papers using the full eligibility checklist to decide if they should be included in preselection. This was done to exclude papers that involved studies inclined toward medical and clinical techniques, for example, studies related to biomarkers, photoplethysmogram, implantable devices, and defibrillators and studies related to algorithms, methods, and techniques. The second author reviewed 10% (170/1700) of the papers at this stage, and agreement was verified across both authors. Where any disagreement was found, the paper in question was reviewed again by both authors and discussed to reach an agreement. Both researchers then met and cross-checked 50% of the final preselection list, discussed inconsistencies, and agreed upon a final list that included 61 papers for potential inclusion.

Each of these papers was further assessed in the final stage of the screening process to check if they applied qualitative methods and included qualitative data. Any paper that contained both quantitative and qualitative data was included in the final review, but only qualitative data in these papers were analyzed. A total of 25 papers were found to include no data, and 20 papers included only quantitative data. These papers were excluded. This left 16 papers that included qualitative data in our final analysis. Figure 1 provides an overview of the full screening process.

Figure 1. Flow diagram illustrating the screening and selection process of papers.
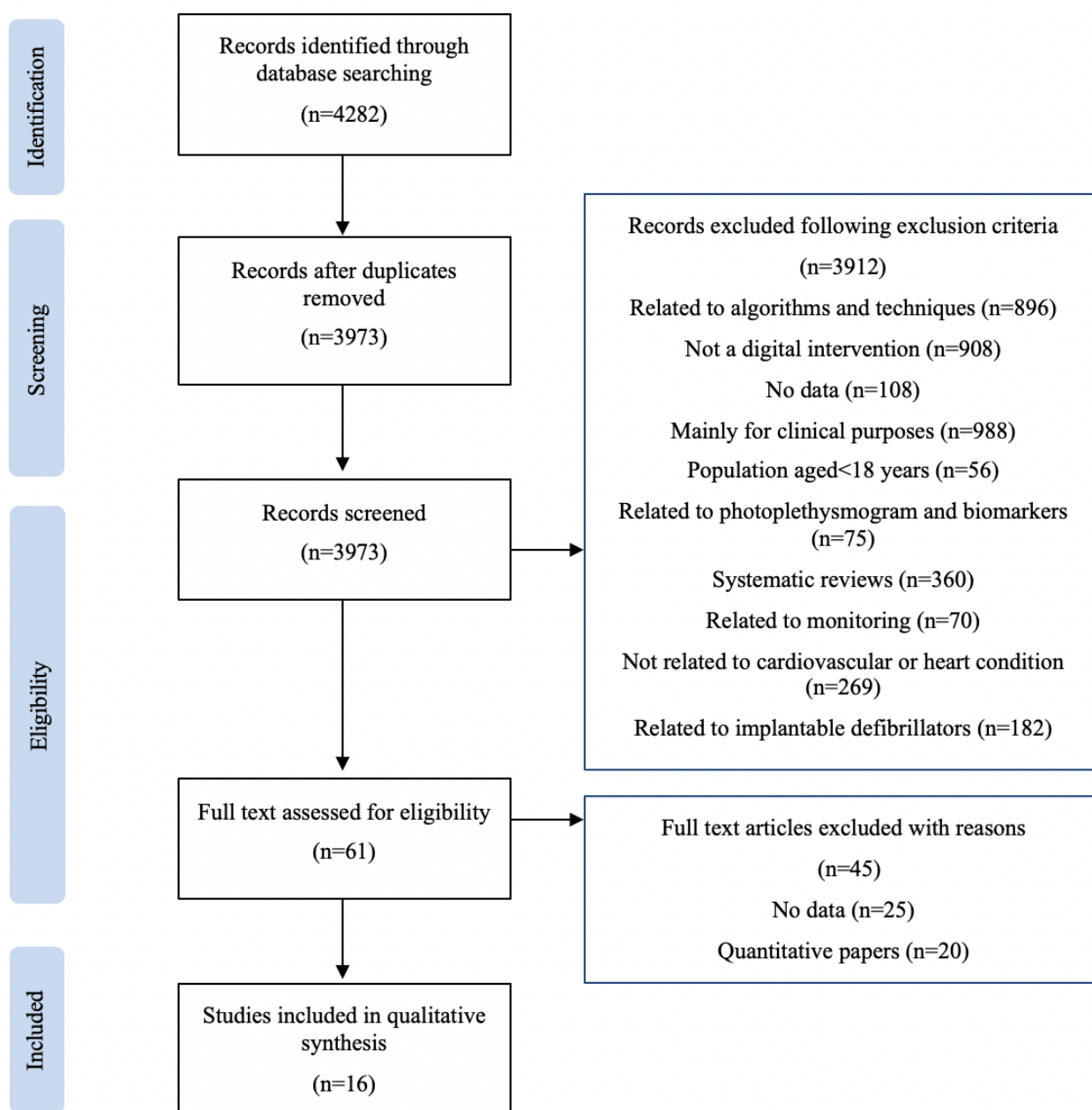
The critical appraisal skills program (CASP) checklist [39] was used to assess the quality of included studies and avoid the risk of bias. The CASP checklists are divided into 3 sections to assess internal validity, results, and relevance to the practice of published papers, and these sections are assessed by questions that can be answered with *yes*, *no*, or *can't tell*. On the basis of the number of questions scored *yes*, an overall rating of *strong*, *moderate*, or *weak* was given to each study. The results of the assessment indicate that the majority of the papers included in the review are strong, whereas others are rated as moderate. Full details of the CASP assessment are included in Multimedia Appendix 3 [33,40-53].

Data from the included papers were initially extracted based on the keywords used in the search terms and eligibility criteria (Textboxes 1 and 2). This included data such as the number of participants, study methods, and settings for each study. In the

final stage of data extraction, the full findings and discussion sections of each of the 16 papers were extracted. This provided data for our subsequent analysis.

Analysis and Synthesis

The analysis step of the GTLR method involves a comparative analysis process with 3 levels of coding: open coding, axial coding, and selective coding. From the set of papers in the final review, ST selected a random paper and carefully read it again, highlighting principal findings, which the GTLR method calls excerpts. Similarly, excerpts from each paper were then listed. At the axial coding stage, these excerpts were articulated to form groups or insights. Both authors carried out an affinity mapping exercise on these excerpts. This led to the formation of groups and subgroups of the excerpts. At the selective coding stage, these groups were then compared and moved around,

followed by discussions among the authors to form themes. This process involved iterative back and forth analysis between the excerpts and groups identified, in which stages were repeated and papers reread until a final consensus was reached. The coding process was supported by Boardthing [54], a web-based notice board software that allows individual and collaborative coding and analysis. The themes were repeatedly discussed and refined among the authors, and the analysis was only complete as the final version of the review documentation was ready.

Results

Study Characteristics

As noted earlier, the keyword search retrieved 4282 articles, of which 16 were included in the final analysis. An overview of the included studies is provided in Multimedia Appendix 4 [40-53,55,56].

Target Users

All studies in the final list focus on patients who had gone through or were going through a cardiac condition. Some of the studies specifically targeted patients diagnosed with heart failure, myocardial infarction, and coronary heart disease. Furthermore, some studies particularly involved participants' postcardiac condition awareness and those who were in their CR phase. Some studies also involved physicians, informal caregivers, nurses, and cardiologists as participants. The papers included studies on both CR [45,46,50,51] and self-management [40-44,47-49,52,53,55,56].

Different Technology Support Provided

In general, the papers in this review investigated mobile or web apps, with some integrating sensors, to manage cardiac conditions. Papers featuring a web-based digital intervention were included [40,41,53]. Some studies used mobile [43,45-47], tablets [48], and a combination of web and mobile systems [42,44,49,50,55,56] as digital interventions. Overall, 2 studies did not involve any particular system. Instead, they focused on patients' needs and perspectives of using an existing technology

and the potential of digital interventions for cardiac management [51,52].

Motivation of the Studies

In general, support for self-management was provided through apps that aim to increase adherence, motivation, and engagement. These could be achieved through gamification [45], by providing guidance and education about the condition [40,43,47,52,55], through reminders and notifications, or by using patient data and sensor data to track and show their progress [46,49,52]. Many studies have involved interventions to increase physical activity and exercise for cardiac patients [41,44,46,48]. Studies also aimed to facilitate better connection between patients and care providers, nurses, or health professionals by providing a medium to communicate and share data [43,53,56]. Two papers were about virtual and remote CR to enable rehabilitation for patients in rural and distant locations [46,50]. One study focused on gathering the needs and interests of patients with CVD to effectively enable remote CR [51].

Design Approaches Used in the Studies

Table 1 provides an overview of the design methods and guiding theories used in the studies. Overall, as all the papers in the final list are qualitative studies, most of the papers used surveys, interviews, and usability tests and represented their evaluation and findings through themes (Table 1). Among these, some studies used theoretical frameworks of behavior change and user-centered design approaches and methodologies. Examples include scenario-based tests, card sorting, goal-directed design, and persuasive design [41-44,47,50,53,55,56]. One study used grounded theory to identify themes from participant responses [49]. Another study used gamification design principles to design the system with the aim of increasing motivation and adherence to lifestyle changes [45]. One study assessed the usability of technology using satisfaction surveys [48], another used a technology usage questionnaire to understand technology usage [51], and another used the system usability scale to assess the usefulness of a system [46].

XSL•FO
RenderX

Table 1. Overview of the theories and design approaches used in the final review.

Study	Design method or guiding theory	Users involvement
Dithmer et al [45]	Gamification and gameful design principles (PERMA[a]) are used to design the app. Gamification principles such as badges, levels, and leader boards were used to increase engagement and motivation.	Requirements gathering, design or prototyping, and evaluation or validation
Yehle et al [49]	No particular design principles or theory and design methodology mentioned.	Requirements gathering and evaluation or validation
Villalba et al [56]	Goal-directed design methodology is applied. A three-phase design process is used: conceptualization, implementation, and validation.	Requirements gathering and evaluation or validation
Jarvis-selinger et al [53]	Diffusion of innovation theory was used as the theoretical lens along with the current telehealth literature for sensitizing concepts. The study used a qualitative methodology, employing a constructivist approach.	Requirements gathering
Fischer et al [40]	Used common sense model of illness representation and showed visualization of body structure and behavior based on different symptoms through a web-based app.	Evaluation or validation
Pfaeffli et al [42]	A library of text and video messages were developed using self-efficacy theory framework and published exercise guidelines.	Requirements gathering, design or prototyping, and evaluation or validation
Katalinic et al [48]	No particular design principles or theory and design methodology mentioned.	Evaluation or validation
Antypas and Wangberg [41]	Different models of health behavior change are combined to form the tailoring algorithm. Tailoring is used as the theoretical framework. A methodological approach that is used to combine the user input and health behavior theory to develop a physical activity digital intervention for cardiac rehabilitation.	Requirements gathering and evaluation or validation
Geurts et al [44]	The prototype design was guided by 3 pillars: simplicity and ease of use, reduce fear and anxiety, and direct and indirect motivation. A human-computer interaction perspective is given by categorizing design decisions according to 3 pillars and show how these pillars resulted in concrete app features.	Requirements gathering, design or prototyping, and evaluation or validation
Buys et al [51]	No particular design principles or theory and design methodology mentioned.	Requirements gathering
Cornet et al [47]	Three frameworks guided the design process: Systems Engineering Initiative for Patient Safety (Version 2.0), Patient Work Framework, and user-centered design.	Requirements gathering, design or prototyping, and evaluation or validation
Banner et al [50]	No particular design principles or theory and design methodology mentioned.	Evaluation or validation
Baek et al [43]	No particular design principles or theory and design methodology mentioned.	Requirements gathering, design or prototyping, and evaluation or validation
Salvi et al [55]	Fogg's Persuasive Systems Design principles were used when designing the GEx system, and health belief models were used to classify patients on the basis of the perceived benefits and barriers to self-efficacy in healthy behavior. The system design and development were guided by a combination of methodologies: Goal-Directed Design, Persuasive Systems Design, and agile software development. The desired behaviors were mapped into specific system's specifications, borrowing concepts from Fogg's Persuasive Systems Design principles.	Requirements gathering and evaluation or validation
Beatty et al [46]	No particular design principles or theory and design methodology mentioned.	Requirements gathering, design or prototyping, and evaluation or validation
Smith et al [52]	No particular design principles or theory and design methodology mentioned.	Requirements gathering

[a]PERMA: Positive emotion, Relationships, Meaning, and Accomplishment.

The 3 main stages of the HCI design process included in the ISO 9241 HCI development lifecycle are requirements gathering, producing design solutions, and evaluating the design against the requirements [57]. There is also a recommendation that this process is iterative, typically involving multiple cycles of design and evaluation. The design process, also known as the user-centered design, focuses on users and their needs in each stage of the process, and iteration continues until it is fit for implementation. We found limited evidence of studies applying a truly iterative approach and user-centered approach. A total of 9 of the 16 papers stated that a user-centered design approach was followed; however, it is not always clear that this involved multiple iterations of the design cycle [41-47,49,56]. Only 6 of the papers provided details of studies that involved users in each stage of the process [42-47]. Moreover, 3 of 16 studies involved users only in the final stage, that is, evaluation [40,48,50].

Users' Perspectives of Digital Interventions for Cardiac Self-Management and Rehabilitation

This section presents the final themes identified in our grounded theory analysis.

Knowledge

Evidence from the review suggests that knowledge plays an important role in rehabilitation and self-management. Education and knowledge influence self-management and increase confidence. To explain this further, we have categorized knowledge into 2 types: background knowledge and personal and in-the-moment understanding.

General Knowledge About CVD

General knowledge or background knowledge about CVD is the fundamental information or awareness that is required to be known by all patients with CVD. This can be information about one's health condition, symptoms, body, medication, preventive measures, and advised lifestyle changes. Background knowledge also includes awareness about different support systems that help people to care for themselves, such as rehabilitation support and digital interventions.

There is a growing trend to use digital interventions to provide the required educational support. A study conducted to validate a self-care digital system to manage cardiovascular condition at home emphasized that education on symptoms and medication was highly valued by patients and health professionals; however, younger patients had reservations about lifestyle education, as they considered it to be intrusive and annoying. Similarly, patients who were initially scared of new technologies, later, after introductory explanations, found it easier to interact with the system [56]. Similarly, a study that evaluated the use of web-based visualizations of patient parameters to improve patients' understanding of their disease and increase their level of control over the rehabilitation process shows that enhanced knowledge and understanding of the illness and its symptoms can motivate protective action, such as for individuals with heart failure to improve self-management of the illness and the symptoms [40]. For example:

> Now I understand why my legs always swelled up. [40]

> We truly know how to, what is happening inside his heart, and why he's getting all these symptoms. In the 2 years that we've been dealing with this illness, it's so good to have it summarised up so that we know how to care for ourselves better. [40]

Participants also repeatedly referred to the need to find the right answers either through an online forum or some kind of knowledge bank [41]:

> It should be a forum where you have the opportunity to get the right answers, access to a resource, this is what I believe it becomes. It has an effect. [41]

CR classes are also popularly known to provide essential knowledge, guidance, and support for patients:

> ...Your class (cardiac rehabilitation) because they stressed what is really bad for you and what is good for you so that makes you stop and think when you are even buying your groceries to make sure you are getting the right stuff. [49]

Personal and In-the-Moment Understanding

Personal and in-the-moment understanding is the supplementary information that patients seek to enhance their self-care process. This type of information is acquired through personal tracking and monitoring and refers to the ongoing knowledge people develop about their individual condition. Knowing one's body plays a key role is achieving control of the cardiac condition; however, it may be difficult to notice some changes and trends in everyday life. Technology has been used to make health and contextual information more easily available to patients and caregivers on an ongoing basis [23]. Patients state that being monitored by technology increases their feeling of security and comfort by enabling a better ongoing understanding of their health [56]. Self-care technologies that use monitored data to guide people to exercise or train within recommended or safe zones boosted confidence and increased motivation:

> The application is not only beneficial for people who are afraid to exercise, but also supports people that have a higher risk to train too much. [44]

A study conducted to understand the current technology usage of patients with CVD and to understand their needs and interests found that ongoing advice on exercise ideas, exercise prompts, information on local exercise opportunities, healthy meal ideas and recipes, and practical ideas to manage stress received the highest ratings for inclusion in a technology-based CR platform [51]:

> I am unsure if I am doing the right thing, like food, so I like advice on that. [52]

Social Versus Individual

Although most patients often manage their care autonomously, clinicians, other people living with the same condition, and caregivers play an equally important role.

Individual Responsibility

Responsibility for change in behavior is personal [41]. Changing behavior is easier if new habits are created by replacing old bad ones. To retain changes, it is important to make it part of the daily routine. Ubiquitous technology can support behavior change in the challenging situations of everyday life and remind users of their own commitments:

> If you could get a message every day, there and then? [41]

> I believe that someone gets used to it, if we make a system, habits. That it doesn't get too much, that we know that...we go online...and we get our own responsibility of our own training. [41]

Technology can support small personal achievements such as getting out of the house to get fit. The use of digital systems as a tool for self-management is valued, especially among the younger ones:

> It gave me the opportunity to get out of the home and try and get myself fit after the operation. I believe it

has achieved that and more. I feel better in myself and I can achieve most jobs without taking about it. [55]

Connecting With Others

Patients often seek to connect with others living with the same condition, and they use these interactions to understand how to live with their condition, validate their assumptions about their body and self-care, and obtain emotional support [58]. A CR session is an excellent example of this type of environment. A theme repeatedly expressed by the patients of the CR program was the importance of not being alone in the rehabilitation and self-management process. This was an important factor that helped them during their visits to the rehabilitation center, and it was something they wished to maintain after their discharge [41]. In addition, CR attenders found great value in being able to ask nurses, cardiologists, and dietitians questions according to their specific needs [42]. Digital interventions also provide easy access to others with the same condition, health professionals, and experts. A study on the experiences of patients undergoing virtual cardiac rehabilitation program (vCRP) demonstrated the potential of vCRP as a medium to provide easy access to health care professionals, nurses, exercise specialists, and dieticians. Although there were some concerns about trust and privacy [41], many of the participants explained that having ongoing monitoring from health care providers as well as support for self-management activities helped them adhere to their recommended program:

> *You know I had stents four years ago, and you start off with the best of intentions, but nobody looks over your shoulder and you peter out. At this time, I felt this is a nifty program...somebody's watching it and I better do it. Keeps you honest, keeps you focused.* [50]

Keeping in touch with the group helps to lift people's mood, is comforting, and provides support; therefore, many patients liked to use forums and web-based groups. Groups and forums on the internet are seen to help individuals be more committed to fitness by sharing goal completions and bragging about it for healthy competition. Forums brought more focus and motivation, as it makes individuals feel obliged to do activities. A study that used gamification for telerehabilitation program of patients with CVD also demonstrates the importance of social and family support, with patients stating that the most important aspect of the game was being able to play with a partner, thus enabling them to deal with rehabilitation as a team:

> *Training diary on the Internet...And also have a group where someone can subscribe to a forum, or have a...to brag...yesterday I walked for an hour and today I have been to the training...and tomorrow I have thought, yes...So, it is like this that someone gets to, a bit, a bit like a competition, internally between each of us. We will train, as much as possible we will commit to ourselves a bit more also.* [41]

> *I am saying that if we have it fixed, one time per week, that we send a message to each other and then, then you feel committed to say yes, for as long as you like...Yes, then you must have something else that*

really, you have something else that you have to do, or else...you just do it. [41]

Motivation and Demotivation

The systems in the listed papers took a number of approaches to provide engagement and motivation toward self-management. Some of the key features of technology and patients' attitudes toward them are described below.

Feedbacks and Reminders

Digital health interventions such as text messages and mobile- and web-based app reminders push patients to maintain the desired changes [42]. Apps using gamification principles are considered motivating, as they allow score, activity and goal comparison, healthy challenges, and competitions. Creating small manageable tasks was positively received by heart patients. Apps use data visualizations to show meaningful comparisons and to see how well they progressed [45]:

> *I went cycling without the application today, but it was less fun!* [44]

> *Two teams explicitly stated that on a day with bad weather, they would not have gone for a walk had they not been motivated by the application.* [45]

Reminders in any form were positively accepted by the patients. Text messages, although intrusive, pushed them to perform exercises, and many stated that reminders such as an alarm are needed for medication management [56]. On the other hand, some patients did not like reminders, as they constantly reminded them of their sickness.

Tracking and Monitoring

Digital health interventions that had the ability to track patients' activities, heart rate, and current health status and showed their progress over time were considered valuable and engaging [44]. In a study to understand the current technology usage of patients with CVD, 68% of patients reported that heart rate monitoring was important when exercising at home [51]. In addition, patients also anticipated that they would be able to manage their disease more efficiently if their daily data could be easily entered in an app and shared with their doctors [43]:

> *I like the fact that I can put all of that and track it, and that my doctors can as well. I can show my doctor what I've been working on.* [46]

> *I think that the idea of an app that records all of the information that this app is doing will be very valuable. Actually somewhat of a motivation for me to do this thing.* [46]

Personalization

Some studies in this review suggested that digital interventions that gave the user the ability to personalize the app based on personal interests contributed toward motivation [41,44]. For example, one of the patients in a study that evaluated patients' motivation when using a mobile app that guided them while cycling suggested that the app would be more engaging and fun if it had the flexibility to insert his preferred routes along with the preloaded ones. However, another patient in the same study preferred predefined routes [44]. Another study showed that

XSL•FO
RenderX

although patients preferred simple interaction methods, they also asked for the possibility of applying advanced settings [51]. The findings of the same study also suggested that the future of technology-enabled CR might include different solutions to reach both men and women to better engage a broader target population of patients with CVD [51].

Increased Burden

Some studies in this review demonstrated patients' concerns regarding using technology. For instance, some patients suggested that adding a device on top of what they already have led to them getting side tracked and thus not using it every day [47]. Patients in the older age group were especially resistant to use technology; some of them lacked interest and found it burdensome:

> I'm retired and I gave all the computerization that I wanted up, that is it I do not even look at it and I will not even turn it on. [53]

Furthermore, lack of time and other priorities is a barrier to self-management and use of technology. Most patients already have measuring devices at home, such as weight scales and blood pressure cuffs, and preferred to continue using devices they already know [56]:

> There are people who like this (application) kind of stuff...and got the time. So for these people it might be great. [47]

Acceptability of Technology

In contrast, studies in this review also demonstrated patients' willingness to use technology. For example, one study reported that patients' interest or intent to use an app for CVD management was high, despite the fact that most were older people who were unfamiliar with the information technology environment [43]. Overall, in most studies, patients as well as clinicians readily accepted and showed interest in learning about new technology [43,48].

Nevertheless, to reach the entire target population of patients with CVD, a variety of technology solutions should be designed to reach both men and women [51].

Usability

Finally, usability and ease of use are crucial for the acceptance of any type of digital intervention and thereby influence engagement. Many studies in this review emphasize that simple interaction methods are preferable. For example, one study stated that 38% of the patients preferred an interaction of no more than a few mouse clicks [51]. Patients unfamiliar with technology positively stated that it was just a matter of getting accustomed, and if they learned and used the app regularly, they would find it simple. Some patients also suggested considering e-literacy issues and initial training [41]:

> It was pretty easy...I like that it's simple. [46]

> I'm not used to this. Once I get used to it, I'll know where everything is. [46]

Discussion

Principal Findings

This review aims to understand users' perspectives of technology in CR and self-management and identify barriers and facilitators of the use of technology. The results suggest that many patients have a positive attitude toward the use of technology. The grounded theory approach enabled us to identify common themes across the included papers, resulting in 3 principal findings:

1. Designers of new technologies and clinicians recommending existing systems to patients should consider seeking the support of both background knowledge and greater in-the-moment understanding. Background knowledge and awareness about the condition and its symptoms, medication, and posthospital care measures are important factors for effective self-management. However, effective self-management also requires patients to be aware of their current body condition and changes in their body, providing reassurance and enabling them to take appropriate measures in self-management.

2. Self-care is a personal responsibility, and people like to try different ways to keep themselves motivated to continue performing self-management activities. For some, but not all, opportunities to stay connected with family, caregivers, and others with a similar health condition are considered as one of the most effective ways to stay motivated and driven toward rehabilitation activities. Again, technology that supports both approaches is likely to be most beneficial.

3. Technologies can use different approaches to support engagement and motivation toward rehabilitation and self-management, including personalization, tracking and monitoring, reminders, and feedback. However, they should take into account the potential to demotivate because of issues including overburdening caused by different devices and apps, privacy concerns, lack of trust, lack of interest, and system usability. If not properly accounted for, these issues can impact the acceptability of systems and become major hindrances to effective rehabilitation and self-management.

These principal findings are discussed in greater detail below and also considered via the lens of relevant HCI literature.

Our first principal finding emphasizes the importance of different types of knowledge. Awareness of available resources, such as awareness of rehabilitation classes, existing online support groups, existing self-care digital apps, and remote rehabilitation videos and programs, is important so that patients can leverage these resources for better and sustained recovery and smoother transition to long-term self-management. In addition, ensuring that patients have knowledge of available emotional and physical support helps to foster self-efficacy if they feel overwhelmed by their CVD condition, leading to the inability to effectivity self-manage [59]. Prior work in HCI has also identified knowledge as an important factor influencing self-care. For example, a study exploring patients' transition from hospitalization to self-management emphasizes gaps in knowledge, resources, and self-efficacy after discharge and

demonstrates an interconnection between them [59]. The study describes knowledge as information provided to patients about their condition, medication, and management and resources as social and physical resources, for example, caregivers and access to health services. Self-efficacy is described as the patient's confidence in their ability to self-manage their condition. The gaps highlighted in that study are consistent with the principal findings of this review. The authors recommend that at a system or hospital level, emphasis on verbal communication of information should be avoided. Ubiquitous computing and embedded technologies could be used to capture and retain verbal information received during hospitalization. In addition, hospitals should provide support and trusted sources of information for patients' access to expertise. On the basis of our findings, these recommendations are also clearly applicable to CR. Similarly, work in HCI describes how patients' understanding of their illness and availability of social and physical resources mediate their self-efficacy [60]. In contrast with prior work, our study has also highlighted the importance of supporting in-the-moment knowledge, which can be acquired through tracking and monitoring. It appears that both types of knowledge can be an integral part of effective CR and self-management.

Effective self-management requires patients to change certain behaviors. An individual's inclination to change behavior depends on the extent to which they are motivated to change [61,62]. Our findings highlight that motivation for action is driven by both individual factors, such as personal responsibility, emotions, and goals, and external influences, such as friends, family, caregivers, health professionals, and personalized and persuasive features of technology. These findings reflect on Deci and Ryan's [61] self-determination theory of motivation, which states that a human's optimal move toward growth is driven by 3 needs: autonomy, the need to have control over one's behavior; relatedness, the need to interact or be connected to others; and competence, the need to experience positive effects of one's activity. Previous HCI research [24,63] provides helpful guidance on how technologies can support these basic needs and also highlights design-related tensions that can arise in balancing different needs. For example, Nunes et al [24] highlighted tensions in the degree of autonomy to be provided to patients, noting that technologies should take into consideration the different levels of autonomy given to the patients for self-care, as it is highly dependent on the disease and the patient's current condition. Although patients are in charge of their health condition, it is important to reflect on the stages or decisions where a clinician's support is needed. Treatment of CVD relies on a combination of medication and lifestyle changes, and there exists an individual difference in the disease management process. Individual differences refer to how people are similar or different in their ways of thinking, feeling, and behaving [64]. This would include patient demographics, situational or contextual changes, and environment. The transtheoretical model of behavior change [65] suggests that effective behavior change could be obtained if personalized feedback with different motivational levels or at different stages of the behavior change process is provided to people. Therefore, it is important to take these differences into account and leverage technology to provide tailored care.

In the case of health care technologies, the one-size-fits-all approach could hamper effective self-care practices [66,67]. Nunes et al [24] also stressed on integrating self-care technologies in everyday lives by prioritizing the lived experiences of patients. This is also emphasized in discussion of *lived informatics* and *design for interweaving* by Rooksby et al [68]. In other words, for health care technologies to be successfully integrated into an individual's life, it is necessary to acknowledge the everyday life of the individual [5]. Moreover, the results of this review demonstrate that patients' adherence to self-management through health care technologies can be improved if technology does not act as a burden in their daily life and is easy to use.

Digital health interventions draw on 2 central domains of study, those originating in health (eg, medicine, biomedical sciences, and psychology) and in technology disciplines (eg, computer science, HCI, and software engineering). This trend is seen in the papers listed in this review. Blandford et al [69] highlighted 7 areas of contrast in practice between technical and health research. They emphasize that skipping over stages of iterative design before investing in large-scale evaluation of digital health technology leads to suboptimally designed solutions. In the HCI community, there is a growing practice of involving end users early on in the design stage and then throughout the full design and evaluation process. In contrast, the studies listed in this review show limited evidence of applying user-centered and iterative design processes. Blandford et al [69] also suggested that failing to learn how the nuances of design affect user interaction and engagement leads to failure in replicating it in different contexts and propagates risk from one design to another. Future research on technology to support CVD should address these limitations. Involving relevant users, in this case, patients, caregivers, and health professionals, in each stage of the design process will help reduce user experience challenges and increase acceptance, leading to more effective digital health interventions. Core to addressing this limitation is appropriate and focused engagement with key patient groups. In this context, although CVD impacts adults across all age groups, it is important to also recognize that CVD and other chronic illnesses are particularly prominent among older populations, and their distinct challenges and complex needs have important implications for the design of such systems [70]. The effectiveness of user-centered design with older adults can be seen in the increasing number of studies involving older populations in the early design stages [71,72].

Limitations

As the aim of this review is to investigate and obtain subjective evidence of the barriers and facilitators of using technology for CR and self-management, only qualitative papers were considered, and review was limited by the analysis of the included studies. The possibility of subjectivity in analyzing the findings is acknowledged, although strategies to limit bias were undertaken through the process of grounded theory analysis and consultation with a second reviewer. In addition, the included studies had varied sample sizes, and the technology was used for different amounts of time in different studies. We acknowledge that this variation could have had an impact on the themes emerging in this review.

Reflective Statement by Authors

This research was conducted in the Republic of Ireland. It is part of the Eastern Corridor Medical Engineering (ECME) collaborative research project, which seeks to improve cardiovascular health with a broad focus on enhancing user-ready sensor technology; improving smart wearables; reducing the complexity of point-of-care diagnostics; and improving smart, clinically relevant monitoring in the assisted living and rehabilitation environments. ECME is a partnership between 5 academic research centers in Northern Ireland, the Republic of Ireland and Scotland, and the Southern Health & Social Care Trust. It involves collaboration between researchers in the medical and technology fields. Both the authors of this paper are based at the Insight Centre for Data Analytics at Dublin when this study was conducted. ST was raised in India and had lived in Dublin for 2 years at the time of the study. She has experience in User Experience design in mobile and assistive technologies. DC has multigenerational roots in Ireland and is an expert in the field of HCI with a focus on the design of digital health technologies. None of the authors have direct lived experience of CVD. This study did not seek to directly address issues such as ethnicity, social and cultural background, and gender, and standard checklists, including the CASP tool, were used to assess the quality of included studies. However, we recognize the potential for bias, both in its own analysis and in the original research papers.

Conclusions

The primary objective of this review was to apply qualitative methods to answer the following research question: What are the primary barriers to and facilitators and trends of digital interventions to support CR and self-management? Our findings show that the use of technology is acceptable to many people undergoing CR and self-management. Although background knowledge is an important facilitator, technology should also support greater ongoing and in-the-moment understanding. Connectedness is valuable, but to avoid becoming a barrier, technology must also respect and enable individual responsibility. Personalization and gamification can also act as facilitators of engagement, but care must be taken to avoid overburdening people. The findings also highlighted the limited use of iterative, user-centered approaches to guide design in this space. Going forward, further application of user-centered and iterative methods represents a significant opportunity.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist.
[PDF File (Adobe PDF File), 409 KB - jmir_v22i11e18025_app1.pdf]

Multimedia Appendix 2
Search strategy.
[PDF File (Adobe PDF File), 74 KB - jmir_v22i11e18025_app2.pdf]

Multimedia Appendix 3
Qualitative assessment table.
[PDF File (Adobe PDF File), 150 KB - jmir_v22i11e18025_app3.pdf]

Multimedia Appendix 4
Table with overview of included studies.
[PDF File (Adobe PDF File), 183 KB - jmir_v22i11e18025_app4.pdf]

References

1. Cardiovascular Diseases (CVDs). World Health Organization. 2017. URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2020-05-11]
2. Findings from the Global Burden of Disease Study 2017. Institute for Health Metrics and Evaluation. 2018. URL: http://www.healthdata.org/sites/default/files/files/policy_report/2019/GBD_2017_Booklet.pdf [accessed 2020-05-11]
3. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics-2018 update: a report from the American Heart association. Circulation 2018 Mar 20;137(12):e67-492. [doi: 10.1161/CIR.0000000000000558] [Medline: 29386200]

XSL•FO
RenderX

4. Burkhart PV, Sabaté E. Adherence to long-term therapies: evidence for action. J Nurs Scholarsh 2003;35(3):207. [Medline: 14562485]

5. Ballegaard S, Hansen T, Kyng M. Healthcare in Everyday Life: Designing Healthcare Services for Daily Life. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Apr 06 Presented at: CHI'08; April 5-10, 2008; Florence, Italy. [doi: 10.1145/1357054.1357336]

6. Cardiac Rehabilitation. British Heart Foundation. URL: https://www.bhf.org.uk/informationsupport/support/practical-support/cardiac-rehabilitation [accessed 2020-05-11]

7. Daly J, Sindone AP, Thompson DR, Hancock K, Chang E, Davidson P. Barriers to participation in and adherence to cardiac rehabilitation programs: a critical literature review. Prog Cardiovasc Nurs 2002;17(1):8-17. [doi: 10.1111/j.0889-7204.2002.00614.x] [Medline: 11872976]

8. Neubeck L, Freedman SB, Clark AM, Briffa T, Bauman A, Redfern J. Participating in cardiac rehabilitation: a systematic review and meta-synthesis of qualitative data. Eur J Prev Cardiol 2012 Jun;19(3):494-503. [doi: 10.1177/1741826711409326] [Medline: 22779092]

9. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]

10. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 2;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]

11. Piette J, List J, Rana G, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015 Nov 24;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]

12. Jörntén-Karlsson M, Pintat S, Molloy-Bland M, Berg S, Ahlqvist M. Patient-centered interventions to improve adherence to statins: a narrative synthesis of systematically identified studies. Drugs 2016 Oct;76(15):1447-1465 [FREE Full text] [doi: 10.1007/s40265-016-0640-x] [Medline: 27677773]

13. Foster EJ, Munoz S, Crabtree D, Leslie SJ, Gorely T. Barriers and facilitators to participating in cardiac rehabilitation and physical activity in a remote and rural population; a cross-sectional survey. Cardiol J 2019 Sep 30:- epub ahead of print [FREE Full text] [doi: 10.5603/CJ.a2019.0091] [Medline: 31565795]

14. Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. Nat Rev Cardiol 2018 Apr;15(4):230-240. [doi: 10.1038/nrcardio.2017.154] [Medline: 29022571]

15. Gallagher R, Roach K, Sadler L, Glinatsis H, Belshaw J, Kirkness A, et al. Mobile technology use across age groups in patients eligible for cardiac rehabilitation: survey study. JMIR Mhealth Uhealth 2017 Oct 24;5(10):e161 [FREE Full text] [doi: 10.2196/mhealth.8352] [Medline: 29066425]

16. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

17. Hekler EB, Buman MP, Poothakandiyil N, Rivera DE, Dzierzewski JM, Morgan AA, et al. Exploring behavioral markers of long-term physical activity maintenance: a case study of system identification modeling within a behavioral intervention. Health Educ Behav 2013 Oct;40(1 Suppl):51S-62S [FREE Full text] [doi: 10.1177/1090198113496787] [Medline: 24084400]

18. Winter SJ, Sheats JL, King AC. The use of behavior change techniques and theory in technologies for cardiovascular disease prevention and treatment in adults: a comprehensive review. Prog Cardiovasc Dis 2016;58(6):605-612 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]

19. Sankaran S, Frederix I, Haesen M, Dendale P, Luyten K, Coninx K. A Grounded Approach for Applying Behavior Change Techniques in Mobile Cardiac Tele-Rehabilitation. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2016 Presented at: PETRA'16; June 29-July 1, 2016; Corfu, Greece p. 1-8. [doi: 10.1145/2910674.2910680]

20. Norman D, Draper S. User Centred System Design: New Perspectives on Human/Computer Interaction. New Jersey, USA: Lawrence Erlbaum Associates Inc; 1986.

21. Dix A, Finlay JE, Abowd GD, Beale R. Human-Computer Interaction. London, England: Pearson Education; 2003.

22. Preece J, Sharp H, Rogers Y. Interaction Design: Beyond Human-Computer Interaction. New Jersey,USA: John Wiley & Sons; 2015.

23. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the User Interface: Strategies for Effective Human-Computer Interaction. London, England: Pearson; 2010.

24. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-care technologies in HCI: trends, tensions, and opportunities. ACM Trans Comput Hum Interact 2015 Dec 14;22(6):1-45. [doi: 10.1145/2803173]

25. de vito Dabbs A, Myers BA, McCurry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]

26. Balaam M, Egglestone S, Fitzpatrick G, Rodden T, Hughes A, Wilkinson A, et al. Motivating Mobility: Designing for Lived Motivation in Stroke Rehabilitation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Mar Presented at: SIGCHI'11; May 7-12, 2011; Vancouver BC p. 3073-3082. [doi: 10.1145/1978942.1979397]

27. Ayoade M, Baillie L. A Novel Knee Rehabilitation System for the Home. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: SIGCHI'14; April 26-May 1, 2014; Toronto, Canada p. 2521-2530. [doi: 10.1145/2556288.2557353]

28. Wolfswinkel JF, Furtmueller E, Wilderom CP. Using grounded theory as a method for rigorously reviewing literature. Eur J Inf 2017 Dec 19;22(1):45-55. [doi: 10.1057/ejis.2011.51]

29. Sankaran S, Frederix I, Haesen M, Dendale P, Luyten K, Coninx K. A Grounded Approach for Applying Behavior Change Techniques in Mobile Cardiac Tele-Rehabilitation. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2016 Presented at: PETRA '16; June 29-July 1, 2016; Corfu, Greece p. 1-8. [doi: 10.1145/2910674.2910680]

30. Mckay E, Askenäs L, Aidemark J. Participatory Learning: Designing for Mobile Enhanced Self-managed Healthcare (MESH). In: Proceedings of the IADIS International Conference ICT, Society and Human Beings. 2015 Jul Presented at: ICT'15; July 21-23, 2015; Las Palmas de Gran Canaria, Spain.

31. Smaradottir B, Martinez S, Borycki E, Loudon G, Kushniruk A, Jortveit J, et al. User evaluation of a smartphone application for anticoagulation therapy. Stud Health Technol Inform 2018;247:466-470. [Medline: 29678004]

32. Puri C, Ukil A, Bandyopadhyay S, Singh R, Pal A, Mandana K. iCarMa: Inexpensive Cardiac Arrhythmia Management: An IoT Healthcare Analytics Solution. In: Proceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems. 2016 Jun Presented at: IoT of Health'16; June 26-30, 2016; Singapore, Singapore p. 3-8. [doi: 10.1145/2933566.2933567]

33. Villalba E, Salvi D, Ottaviano M, Peinado I, Arredondo M, Akay A. Wearable and mobile system to manage remotely heart failure. IEEE Trans Inf Technol Biomed 2009 Nov;13(6):990-996. [doi: 10.1109/TITB.2009.2026572] [Medline: 19643715]

34. Andersen T, Andersen P, Kornum A, Larsen T. Understanding Patient Experience: A Deployment Study in Cardiac Remote Monitoring. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 May Presented at: PervasiveHealth'17; May 23-26, 2017; Barcelona, Spain p. 221-230. [doi: 10.1145/3154862.3154868]

35. Grimes A, Kantroo V, Grinter R. Let's Play! Mobile Health Games for Adults. In: Proceedings of the 12th ACM international conference on Ubiquitous computing. 2010 Presented at: ACM'10; September 26-29, 2010; Copenhagen, Denmark p. 241-250. [doi: 10.1145/1864349.1864370]

36. Brandt C, Clemensen J, Nielsen J, Søndergaard J. Drivers for successful long-term lifestyle change, the role of e-health: a qualitative interview study. BMJ Open 2018 Mar 12;8(3):e017466 [FREE Full text] [doi: 10.1136/bmjopen-2017-017466] [Medline: 29530904]

37. Choe E, Lee N, Lee B, Pratt W, Kientz J. Understanding Quantified-selfers' Practices in Collecting and Exploring Personal Data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI'14; April 26-May 1, 2014; Toronto, Canada p. 1143-1152. [doi: 10.1145/2556288.2557372]

38. Consolvo S, McDonald D, Toscos T, Chen M, Froehlich J, Harrison B, et al. Activity Sensing in the Wild: a Field Trial of Ubifit Garden. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: ACM CHI'08; April 5-10, 2008; Florence, Italy p. 1797-1806.

39. Critical Appraisal Skills Programme. URL: https://casp-uk.net/ [accessed 2020-06-12]

40. Fischer S, Wünsche B, Cameron L, Morunga E, Parikh U, Jago L, et al. Web-based Visualisations Supporting Rehabilitation of Heart Failure Patients by Promoting Behavioural Change. In: Proceedings of the Thirty-Fourth Australasian Computer Science Conference - Volume 113. 2011 Presented at: ACSC'11; January 17-20, 2011; Perth, Australia p. 53-62 URL: http://dl.acm.org/citation.cfm?id=2459296.2459303 [doi: 10.5555/2459296.2459303]

41. Antypas K, Wangberg S. Combining users' needs with health behavior models in designing an internet- and mobile-based intervention for physical activity in cardiac rehabilitation. JMIR Res Protoc 2014 Jan 10;3(1):e4 [FREE Full text] [doi: 10.2196/resprot.2725] [Medline: 24413185]

42. Pfaeffli L, Maddison R, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mHealth cardiac rehabilitation exercise intervention: findings from content development studies. BMC Cardiovasc Disord 2012 May 30;12:36 [FREE Full text] [doi: 10.1186/1471-2261-12-36] [Medline: 22646848]

43. Baek H, Suh J, Kang S, Kang S, Lim TH, Hwang H, et al. Enhancing user experience through user study: design of an mHealth tool for self-management and care engagement of cardiovascular disease patients. JMIR Cardio 2018 Feb 9;2(1):e3 [FREE Full text] [doi: 10.2196/cardio.9000] [Medline: 31758783]

44. Geurts E, Haesen M, Dendale P, Luyten K, Coninx K. Back on Bike: the BoB Mobile Cycling App for Secondary Prevention in Cardiac Patients. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2016 Presented at: MobileHCI '16; September 6-9, 2016; Florence, Italy p. 135-146. [doi: 10.1145/2935334.2935377]

45. Dithmer M, Rasmussen J, Grönvall E, Spindler H, Hansen J, Nielsen G, et al. 'The Heart game': using gamification as part of a telerehabilitation program for heart patients. Games Health J 2016 Feb;5(1):27-33 [FREE Full text] [doi: 10.1089/g4h.2015.0001] [Medline: 26579590]

46. Beatty A, Magnusson S, Fortney J, Sayre G, Whooley M. VA FitHeart, a mobile app for cardiac rehabilitation: usability study. JMIR Hum Factors 2018 Jan 15;5(1):e3. [doi: 10.2196/humanfactors.8017] [Medline: 29335235]

47. Cornet VP, Daley CN, Srinivas P, Holden RJ. User-centered evaluations with older adults: testing the usability of a mobile health system for heart failure self-management. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):6-10 [FREE Full text] [doi: 10.1177/1541931213601497] [Medline: 30930610]

48. Katalinic O, Young A, Doolan D. Case study: the interact home telehealth project. J Telemed Telecare 2013 Oct;19(7):418-424. [doi: 10.1177/1357633X13506513] [Medline: 24218357]

49. Yehle K, Chen A, Plake K, Yi J, Mobley A. A qualitative analysis of coronary heart disease patient views of dietary adherence and web-based and mobile-based nutrition tools. J Cardiopulm Rehabil Prev 2012;32(4):203-209 [FREE Full text] [doi: 10.1097/HCR.0b013e31825b4e6a] [Medline: 22760245]

50. Banner D, Lear S, Kandola D, Singer J, Horvat D, Bates J, et al. The experiences of patients undertaking a 'virtual' cardiac rehabilitation program. Stud Health Technol Inform 2015;209:9-14. [Medline: 25980699]

51. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Jul 19;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]

52. Smith R, Menon J, Rajeev J, Feinberg L, Kumar R, Banerjee A. Potential for the use of mHealth in the management of cardiovascular disease in Kerala: a qualitative study. BMJ Open 2015 Nov 17;5(11):e009367 [FREE Full text] [doi: 10.1136/bmjopen-2015-009367] [Medline: 26576813]

53. Jarvis-Selinger S, Bates J, Araki Y, Lear SA. Internet-based support for cardiovascular disease management. Int J Telemed Appl 2011;2011:342582. [doi: 10.1155/2011/342582] [Medline: 21822430]

54. Boardthing. 2017. URL: https://boardthing.com/ [accessed 2019-10-11]

55. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]

56. Villalba E, Peinado I, Arredondo M. Self Care System to Assess Cardiovascular Diseases at Home. In: International Conference on Universal Access in Human-Computer Interaction. 2009 Presented at: UAHCI'09; July 19-24, 2009; San Diego, California, USA p. 248-257. [doi: 10.1007/978-3-642-02710-9_28]

57. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. In: User Centered System Design: New Perspectives on Human-computer Interaction. Boca Raton, Florida: CRC Press; 2020.

58. Pols J. Care at a Distance: On the Closeness of Technology. Amsterdam, Netherlands: Amsterdam University Press; 2012:1.

59. Pollack AH, Backonja U, Miller AD, Mishra SR, Khelifi M, Kendall L, et al. Closing the gap: supporting patients' transition to self-management after hospitalization. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:5324-5336 [FREE Full text] [doi: 10.1145/2858036.2858240] [Medline: 27500285]

60. Di EB, Cameron R, Boyd A, Lopez K, Martyn-Nemeth P, Dickens C, et al. Designing Self-care Technologies for HF Patients: a Conceptual Model. In: Conference on Human Factors in Computing Systems. 2019 Mar Presented at: CHI'19; March 12-16, 2019; New York, USA.

61. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]

62. Deci E, Ryan R. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. Psychological inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]

63. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018 May;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]

64. Chen Y. Take It Personally: Accounting for Individual Difference in Designing Diabetes Management Systems. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems. 2010 Presented at: ACM'10; April 1-4, 2010; Washington, DC p. 252-261. [doi: 10.1145/1858171.1858218]

65. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]

66. Ferron M, Massa P. Transtheoretical Model for Designing Technologies Supporting an Active Lifestyle. In: Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI. 2013 Sep Presented at: CHI'13; September 16-20, 2013; Trento, Italy. [doi: 10.1145/2499149.2499158]

67. Storni C, Bannon L. Towards the design of truly patient-centred healthcare infrastructures: a socio-technical approach to self-care. Heal Informatics An Adapt Commun Technol Futur Healthc 2012.

68. Rooksby J, Rost M, Morrison A, Chalmers M. Personal Tracking as Lived Informatics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, Canada p. 1163-1172. [doi: 10.1145/2556288.2557039]

69. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]

70. Jochems N. Designing Tablet Computers for the Elderly A User-Centered Design Approach. In: International Conference on Human Aspects of IT for the Aged Population. 2016 Presented at: ITAP'16; July 17-21, 2016; Toronto, Ontario, Canada p. 42-51. [doi: 10.1007/978-3-319-39943-0_5]

71. Doyle J, Murphy E, Hannigan C, Smith S, Bettencourt-Silva J, Dinsmore J. Designing Digital Goal Support Systems for Multimorbidity Self-Management: Insights from Older Adults and their Care Network. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18; May 21-24, 2018; New York City, USA p. 168-177. [doi: 10.1145/3240925.3240982]

72. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 9;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]

Abbreviations

CASP: critical appraisal skills program
CR: cardiac rehabilitation
CVDs: cardiovascular diseases
ECME: Eastern Corridor Medical Engineering
EU: European Union
GTLR: grounded theory literature review
HCI: human-computer interaction
MeSH: Medial Subject Headings
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
vCRP: virtual cardiac rehabilitation program

Review

Mindfulness-Based Programs for Patients With Cancer via eHealth and Mobile Health: Systematic Review and Synthesis of Quantitative Research

Juraj Matis[1], MA; Miroslav Svetlak[2,3], PhD; Alena Slezackova[1,2], PhD; Marek Svoboda[4], PhD; Rastislav Šumec[2,5], PhD

[1]Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic

[2]Department of Psychology and Psychosomatics, Faculty of Medicine, Masaryk University, Brno, Czech Republic

[3]Department of Psychiatry, Faculty of Medicine, Masaryk University, Brno, Czech Republic

[4]Department of Comprehensive Cancer Care, Masaryk Memorial Cancer Institute, Brno, Czech Republic

[5]First Department of Neurology, Faculty of Medicine, Masaryk University and St. Anne's University Hospital, Brno, Czech Republic

Corresponding Author:
Miroslav Svetlak, PhD
Department of Psychology and Psychosomatics
Faculty of Medicine
Masaryk University
Kamenice 126/3
Brno, +42062500
Czech Republic
Phone: 420 773111330
Email: msvetlak@med.muni.cz

Abstract

Background: eHealth mindfulness-based programs (eMBPs) are on the rise in complex oncology and palliative care. However, we are still at the beginning of answering the questions of how effective eMBPs are and for whom, and what kinds of delivery modes are the most efficient.

Objective: This systematic review aims to examine the feasibility and efficacy of eMBPs in improving the mental health and well-being of patients with cancer, to describe intervention characteristics and delivery modes of these programs, and to summarize the results of the included studies in terms of moderators, mediators, and predictors of efficacy, adherence, and attrition.

Methods: In total, 4 databases (PubMed, PsycINFO, Scopus, and Web of Knowledge) were searched using relevant search terms (eg, mindfulness, program, eHealth, neoplasm) and their variations. No restrictions were imposed on language or publication type. The results of the efficacy of eMBPs were synthesized through the summarizing effect estimates method.

Results: A total of 29 published papers describing 24 original studies were included in this review. In general, the results indicate that eMBPs have the potential to reduce the levels of stress, anxiety, depression, fatigue, sleep problems, and pain, and improve the levels of mindfulness, posttraumatic growth, and some parameters of general health. The largest median of Cohen d effect sizes were observed in reducing anxiety and depression (within-subject: median –0.38, IQR –0.62 to –0.27; between-group: median –0.42, IQR –0.58 to –0.22) and facilitating posttraumatic growth (within-subject: median 0.42, IQR 0.35 to 0.48; between-group: median 0.32, IQR 0.22 to 0.39). The efficacy of eMBP may be comparable with that of parallel, face-to-face MBPs in some cases. All studies that evaluated the feasibility of eMBPs reported that they are feasible for patients with cancer. Potential moderators, mediators, and predictors of the efficacy, attrition, and adherence of eMBPs are discussed.

Conclusions: Although the effects of the reviewed studies were highly heterogeneous, the review provides evidence that eMBPs are an appropriate way for mindfulness practice to be delivered to patients with cancer. Thus far, existing eMBPs have mostly attempted to convert proven face-to-face mindfulness programs to the eHealth mode. They have not yet fully exploited the potential of eHealth technology.

XSL•FO
RenderX

Introduction

Background

Cancer is the second leading cause of death in the world, and approximately 1 in 6 deaths is because of this disease [1]. Owing to medical care, the life of patients has been increasingly prolonged for some types of cancer and associated somatic symptoms such as pain, fatigue, and nausea are better controlled. In this context, health care professionals face the challenge of helping more patients with cancer than ever before to live their lives more fully. Psychosocial distress associated with life for patients with cancer has been identified as a significant problem. In a large group of new patients with cancer (3035 patients), 25.7% scored above the cutoff points for distress, anxiety, and depression [2]. The prevalence of major depressive disorders is approximately 15% in patients with advanced cancer [3], and 30% to 40% of patients in various stages of cancer development report significant psychosocial distress symptoms, such as anxiety, depression, nervousness, and insomnia [4-6]. The management of these psychiatric symptoms, especially of distress related to the cancer diagnosis, is one of the main challenges of complex oncological and palliative care.

Psychotherapy, counseling, and other nonpharmacological methods such as mindfulness-based programs (MBPs) are often not implemented, despite their efficacy in standard oncological and palliative care [7,8]. They are underestimated by patients [9] and physicians [10], and they are often unavailable at the appropriate time.

MBPs

The *first generation* of MBPs, which have the most robust evidence within the field, are mindfulness-based stress reduction (MBSR) and mindfulness-based cognitive therapy (MBCT). The difference between MBSR and MBCT depends partly on the client group at which the course is aimed. MBSR was originally aimed at people with chronic pain and stress [11]. MBCT was aimed at people with an affective disorder, especially those with recurrent depression [12]. MBCT combines systematic mindfulness training with cognitive behavioral therapy to help people with a history of depression learn vital skills. Both involve a manualized 8-week program of meditation and gentle Hatha yoga training. Participants attend weekly group sessions where they are introduced to formal meditation practices, gentle yoga, and psychosocial education. The program also includes a silent meditation retreat day that falls in the second half of the course. Group members are asked to practice for 45 min per day. Participants are encouraged to keep a diary to describe their practice, reflections, and insights. Formal meditation practices include focused attention on breathing, body scans, and open monitoring of sounds, thoughts, feelings, and bodily sensations. Group sessions generally last for 2.5 hours and focus on group meditation practice and discussion of these practices. MBCT is largely based on MBSR, and many of its parts are the same [13].

A growing number of theoretical studies have attempted to operationalize the mindfulness concept. In their oft-cited review, Bishop et al [14] define mindfulness as follows: "Broadly conceptualized, mindfulness has been described as a kind of nonelaborative, nonjudgmental, present-centered awareness in which each thought, feeling, or sensation that arises in the attentional field is acknowledged and accepted as it is." They further state that, "in the state of mindfulness, thoughts and feelings are observed as events in mind, without overidentifying with them and without reacting to them in an automatic, habitual pattern of reactivity. This dispassionate state of self-observation is thought to introduce a 'space' between one's perception and response. Thus, mindfulness is thought to enable one to respond to situations more reflectively (as opposed to reflexively)."

From the first published article by Kabat-Zinn [11] on the positive effect of MBSR on reducing pain and symptoms of negative mood in a group of patients with chronic pain, the effectiveness of MBPs on improving mental and physical health has been repeatedly documented in healthy people [15,16] and in people with various psychiatric or somatic conditions [17,18]. Numerous subsequent studies have also offered considerable evidence about the benefits of practicing mindfulness meditation for patients facing different types and stages of cancer [19-22]. Recent systematic reviews and meta-analyses have documented the moderate positive effect of MBP on anxiety and depression symptoms in patients with cancer and survivors of cancer [23]; a small effect on depression and a moderate effect on anxiety [24]; moderate-to-large positive effects on the mental health of patients with breast cancer [25]; a moderate effect on anxiety, stress, fatigue, general mood, and sleep disturbance and a small effect on physical health variables in a mix of cancer diagnoses [26]; a medium effect on anxiety, depression, quality of life, fatigue, stress, and posttraumatic growth [27]; and a small-to-medium effect on health-related quality of life, fatigue, sleep, stress, anxiety, and depression [28].

MBPs have been repeatedly shown to be effective in reducing cancer-related pain [29], supporting psychological well-being in adults with advanced cancer [30], reducing depressive symptoms in patients with breast cancer [31], decreasing fear of cancer recurrence [32], and even maintaining telomere length in survivors of breast cancer [33]. Systematic and comprehensive programs based on MBCT for cancer (MBCT-Ca) [34] and mindfulness-based cancer recovery (MBCR) [35] have also been developed.

eHealth MBPs

There has been an increasing effort to transfer traditional health care practices to the formats of eHealth and mobile health (mHealth) [36] to provide widely accessible psychological support with minimal economic costs to those who need it.

According to Eysenbach [37], eHealth refers to "health services and information delivered or enhanced through the internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked,

global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology." Later, the term was broadened to include mHealth, adding mobile phones and apps to the definition [36].

Although the therapist-client relationship and dialog are irreplaceable and are key common factors in psychotherapy beyond the effect [38], web-based interventions offer some advantages, specifically for people who are functioning at good or adequate personality levels. Web-based programs (1) are easily accessible, (2) are anonymous, (3) are available 24/7 to people during the course of their daily life, (4) do not necessarily require the involvement of a therapist educated in mindfulness, (5) are less expensive, and (6) save time [39-41]. The preference for web-based delivery is reflected in the increasing number of mindfulness-based mobile apps; a search identified 560 available apps [42].

Increasing evidence supports the advantages and effectiveness of web-based MBPs [41,43,44]. eHealth MBPs (eMBPs) have been shown to be effective in supporting mental health and reducing symptoms of psychopathology in healthy subjects [43,45] and in patients with depression [46], anxiety [47], tinnitus [48], chronic pain [49], and fibromyalgia [50].

In total, 2 recent systematic reviews have documented that eMBPs are feasible and effective for people with various physical health conditions [51] and that the eMBPs are effective in reducing depression and anxiety in clinical populations [52].

An increasing number of studies have revealed that eMBPs are effective and suitable for patients with cancer [53,54]. However, there has not yet been a systematic review of these studies and their descriptions of the interventions in terms of their characteristics, such as delivery mode and approach.

Owing to the heterogeneity in form and content of the eMBPs used in oncology, it is necessary to clarify what an MBP is and what it is not. MBPs typically include mindfulness training via 3 formal mindfulness meditation practices [8]: body scan, mindful movement, and sitting meditation. It is based on daily home practice with the support of recorded guidance. Throughout the program, participants are encouraged to develop their informal practice by bringing awareness in particular ways to everyday life.

According to a study by Crane et al [8], MBPs primarily include MBSR and MBCT, the *first-generation* MBPs. The length, type, and frequency of mindfulness practice are strictly recommended in these programs.

However, as the field has developed, new adaptations of first-generation MBPs have been developed for particular purposes and populations (eg, MBCT-Ca [55] and MBCR [35]). Various adapted MBPs maintain the structure of the first-generation programs and contain the 3 formal mindfulness meditation practices, but they can vary in length and content (eg, with and without a 1-day retreat).

The development in this field has also brought a few approaches, such as acceptance and commitment therapy (ACT) [56] and dialectical behavioral therapy [57] that share several underpinning theoretical ideas with MBPs and some mindfulness

meditation practices. These programs could be called *mindfulness-informed programs* [8]. With ACT, there is some promising evidence that it may improve the quality of life, emotional state, psychological flexibility, and possibly disease self-management in patients with cancer [58,59].

Research on eMBP feasibility and efficacy is still in its initial stages in the oncology field. With the aim of describing the heterogeneity of this field and not missing any potentially feasible and effective eMBPs using the mindfulness approach for patients with cancer, every level of MBP is included in this systematic review: first-generation MBPs, adapted MBPs, and mindfulness-informed programs.

The question of the appropriate length and content for an eMBP to have a positive effect on mental health remains open. MBSR and MBCT programs can be considered as the gold standard in this intervention area with an 8-week duration. Some experimental data have revealed that 4-week mindfulness programs seem to be efficacious for promoting well-being and stress reduction [60]. There is also some consensus among experts in this field [61,62] that a 4-week mindfulness training completion can be considered as a minimum adequate *dose*. The prevalent duration of eMBPs listed in 4 recent systematic reviews [41,51,52,60] is, on average, 8 (SD 1.86) weeks (minimum 3 weeks and maximum 12 weeks). The feasibility and effectiveness of short eMBPs have been demonstrated in patients with cancer undergoing chemotherapy [63], but there are few such programs, and they cannot be considered as systematic mindfulness training.

Aims of This Study

This study aims to examine the feasibility and efficacy of eMBPs in improving mental health and well-being in patients with cancer, to describe the intervention characteristics and delivery modes of these programs, and to summarize the results of the included studies in terms of the moderators, mediators, and predictors of efficacy, adherence, and attrition. This should serve as a starting point for maximizing effectiveness and adherence and minimizing attrition rates in the construction and development of future eMBPs for patients with cancer.

Methods

Reporting Guidelines Used

This review was conducted in accordance with the 2009 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [64] and with supportive guidance from the Cochrane Handbook for Systematic Reviews of Interventions [65]. The protocol of this review was not preregistered.

Search Strategy

A systematic literature search was conducted in 4 electronic databases: PubMed, Web of Knowledge, Scopus, and PsycINFO. Each database was searched from the first available date until July 31, 2020, using relevant search terms (eg, mindfulness, program, eHealth, and neoplasm) and their variations (all terms used are presented in Multimedia Appendix 1). No restrictions were imposed on language or publication type. The World Health Organization, International Clinical Trials Registry

Platform, and the US National Library of Medicine trial registry platform were also searched to detect relevant completed trials that have not yet been published (4 potential trials were found and authors were contacted, but data were not obtained). In addition, the reference lists of the included publications were examined. Figure 1 shows the flowchart of selection and inclusion.

Figure 1. Search and selection process and reasons for exclusion according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines. mHealth: mobile health; ICTRP: International Clinical Trials Registry Platform.
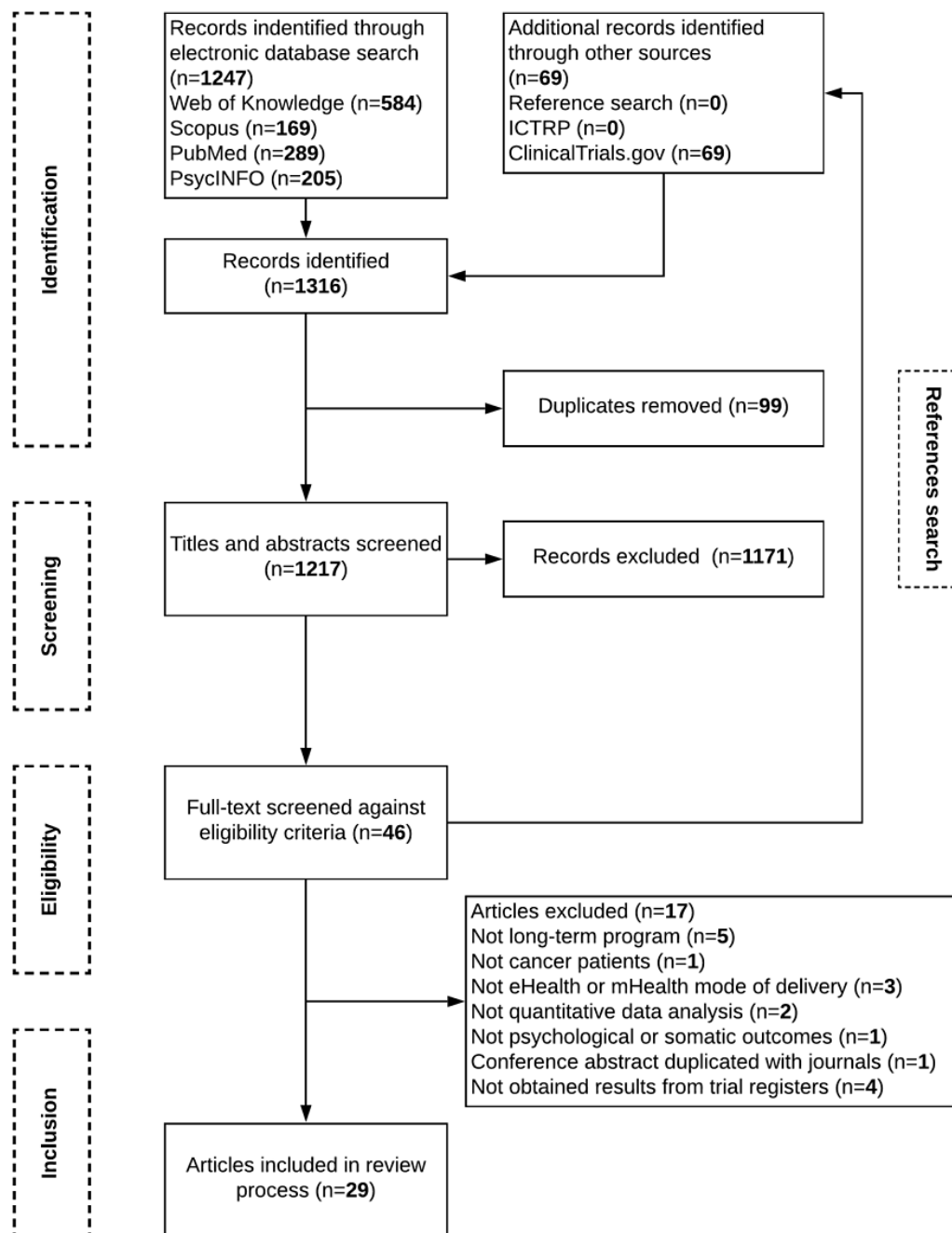
Selection of Studies

Search terms for the literature search were chosen by 2 authors (JM and MS) and then consulted with the review team. The database search and paper screening (title, abstract, and full text) were undertaken by the same 2 authors (JM and MS); if the eligibility of a study was unclear, the review team discussed it until consensus was reached. One study author was contacted

for additional eligibility information; the author responded and supplied the necessary information.

Inclusion criteria were employment of mindfulness in an intervention program, administration of the program via an eHealth mode of delivery (including website, app, videoconference, computer, and telephone), use of quantitative data analysis, evaluation of psychological or somatic outcomes,

administration of the program to a patient population with cancer, and program duration of at least 4 weeks.

Data Extraction

Data extraction was performed by 4 review authors in the first phase (AS, MS, RS, and JM). In the second phase, the extracted data were checked by JM. Disagreements among the review authors were discussed by the review team until consensus was reached. A data extraction sheet was developed and pilot tested on 6 randomly selected included studies and then refined accordingly. Every team member was pretrained in the data extraction process. For each included study, the following data were extracted: first author; country and year of publication; population characteristics, including cancer stage and type, receipt of primary treatment, age, sex (proportion of females in total study population), and number of participants per condition; intervention characteristics, including type (eg, MBSR) and purity (pure mindfulness or combined with other interventions) of intervention, delivery mode (eg, website, telephone), program structure (predefined or nonpredefined; predefined program progress according to guidelines, eg, MBSR; or nonpredefined program progress, eg, free access to program modules based on patient preferences), facilitation (facilitated: synchronous or asynchronous personal contact with the facilitator; nonfacilitated: eMBP without personal facilitation), type and frequency of reminders, presence of a retreat day, number and average time of sessions, and duration of intervention in weeks; comparison group (eg, waitlist, usual care); outcomes and their type (primary or secondary); outcome measurements, duration from baseline to postintervention, or latest available follow-up assessment (eg, 8 weeks); attrition (ie, intervention dropout); adherence to program in terms of session attendance or completion, practice frequency, and practice time or duration; and data to calculate pre-post effect sizes per condition. A total of 3 authors had to be contacted for additional information or missing data; 2 responded and provided the required information or data, and one was not contactable (email out of service).

Evaluation of Methodological Quality

The methodological quality of randomized controlled trial (RCT) studies was assessed based on potential sources of bias outlined in the Cochrane Handbook for Systematic Reviews of Intervention [65]. In pre-post studies, only those items of the Cochrane assessment tool that fit were used. The presence of a control group item was added because almost half of all included studies did not have a control group. Sources of bias that were assessed in all studies (pre-post studies without control group included) were (1) complete outcome data or intention-to-treat (ITT) analysis used (where the threshold for acceptable dropout rate was determined as ≤15%; note that the 10% attrition cutoff recommended by the Cochrane risk of bias tool [66] was modified and set at 15% in this review as a mean attrition rate of eHealth interventions recently systematically reviewed [51]. This cutoff better reflects the higher attrition rate in the eHealth interventions (and it is slightly less conservative), (2) all outcomes reported, and (3) the presence of a control group. Sources of bias that were also assessed only in RCT studies were random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome

assessment (note that because of the nature of the included studies, blinding of participants, personnel, and outcome assessment was not possible), and similar groups (ie, whether the groups were similar on prognostic indicators at baseline or appropriate adjustments were made to correct for baseline imbalance). Note that studies were coded as *yes* when they met the criteria, *no* when they did not meet the criteria, and *unclear* when it was ambiguous as to whether they met the criteria. This results in a low, high, and unclear risk of bias, respectively.

Assessment of methodological quality was undertaken by 4 review authors in the first phase (AS, MS, RS, and JM), and data were checked by JM in the second phase. Disagreements among the authors were discussed by the review team until consensus was reached.

Data Analysis

The adherence rate was computed as the reported amount of practice or sessions (completed by a certain proportion of participants) divided by the intended or recommended amount of the intervention protocol where it was possible. The attrition rate (intervention dropout) was computed as participants who completed the intervention (continued in intervention to the end, regardless of adherence or completion of postintervention assessments) divided by participants enrolled in the intervention. The effect sizes of studies were reported only at the time points when statistical analysis was conducted. When the effect sizes were not reported by the authors and data for the computation were available, they were computed using Cohen d formula [67]. For within-subject results, the formula was as follows: pre-post mean change divided by the pooled SDs. For between-group results, the formula was as follows: the difference between groups pre-post mean changes divided by the pooled baseline SDs.

The efficacy of eMBPs was synthesized using the summarizing effect estimates method. This approach was used because the investigation of eMBPs in patients with cancer is still in its early stages (predominantly in stage 1 [68] according to the National Institutes of Health Stage Model [69], as nearly half of the programs in this review included intervention generation, refinement, modification, adaptation, and pilot testing), whereas efficacy results are reported only in a preliminary manner. Thus, there is a small number of RCT studies per outcome and generally high heterogeneity across all included studies—study characteristics (design, intervention type, delivery modes, measured outcomes, and assessment timepoints) varied substantially, and homogeneity in effects cannot be expected with standard meta-analytical methods. With the aim of identifying potential predictors, moderators, and mediators of eMBP attrition, adherence, and efficacy in the reviewed studies, the results of their analysis are summarized within this review. Possible publication biases could not be estimated because of the limited number of studies per outcome.

Results

Study Selection

The database search and the search of trial registers provided a total of 1316 results. After removing duplicates and screening

titles, abstracts, and full texts against the inclusion criteria, conducting a complementary hand search of the reference lists of eligible studies, and contacting study authors, a total of 29 published papers describing 24 original studies were included in this review. Of the included studies, 5 studies described long-term follow-up results, predictors, and associations or other health-related outcomes that were separately published [53,70-73]. Figure 1 illustrates the search and selection process and reasons for exclusion according to the PRISMA guidelines [64].

Population Characteristics

The population characteristics of the included studies are reported in Multimedia Appendix 2 [74]. Most of the studies examined mixed cancer types (n=11), mixed staged patients (n=11), included patients after the completion of primary cancer treatment (n=11), and were conducted in the United States (n=15). Cancer stages ranged from 0 to 4. Note that if patients received hormonal therapy only, we considered them as having completed primary cancer treatment.

Study Characteristics

The study characteristics of the included studies are reported in Multimedia Appendix 2. A total of 2522 adults participated in the study. Of the included studies, 10 were pre-post pilot feasibility studies without an active or waitlist control group [54,75-83] and 14 employed an RCT design [62,71,84-95]. Of those RCTs, 9 studies had a no-intervention control (waitlist usual care group) [62,71,87-89,91,93-95]; two of those 9 studies used additional active specific intervention–face-to-face MBCT program [62] and supportive-expressive group [89]; 4 studies

compared eMBP with minimal intervention control (education or enhanced usual care) [84-86,90], one of those 4 studies used an additional specific active intervention–physiotherapist-guided ambulant activity feedback (AAF) therapy encompassing the use of an accelerometer [84], and 1 study compared eMBP with a parallel mindfulness program only [92]. Of the included studies, 10 studies reported follow-up results, with the duration from the postintervention assessment ranging from 4 weeks to 9 months.

Intervention Characteristics

The intervention characteristics of the included studies are reported in Multimedia Appendix 2. The primary delivery platform, secondary intervention delivery channel, and type of reminder are summarized in Table 1. The primary platform is an internet or internet-related technology that delivers eMBP. The secondary intervention delivery channel is the concrete way in which the program is delivered to patients. The type of reminder includes the ways in which the patients are reminded of the program in an attempt to increase the patient's completion rate and adherence to the intervention. The most frequently used primary platforms are websites (n=9) and smartphone apps (n=8). All studies except one [71] used audio recordings as a secondary delivery mode (n=23), usually combined with other channels. Reminders were used in almost half of the studies (n=11), and almost half of them used email (n=5). Of all 24 studies, only 2 used the exact same combination of delivery modes; these 2 studies were conducted by the same research group [77,84]. These results reveal considerable heterogeneity in the primary and secondary delivery modes of the included studies.

Table 1. Primary delivery platforms, secondary intervention delivery channels and reminders of eHealth mindfulness-based program of included studies.

Platform	Value, n (%)
Delivery platforms	
Website	9 (38)
Mobile app	8 (33)
Videoconference	5 (21)
Telephone call	3 (13)
Email	2 (8)
Delivery channels	
Audiorecording	23 (96)
Video	10 (42)
Workbook	8 (33)
Email	2 (8)
Reminders	
Email	5 (45)
Telephone call	2 (18)
Text message	1 (9)
Postcard	1 (9)
Notification	1 (9)

The eMBPs could be classified according to program structure (predefined program progress according to guidelines, eg, MBSR; nonpredefined program progress, eg, free access to program modules based on patient preferences) and facilitation (facilitated: synchronous or asynchronous personal contact with the facilitator; nonfacilitated: standardized eMBP without personal facilitation).

A total of 4 eMBPs [54,82,87,93] did not have predefined sessions. They were delivered through mobile apps with various exercises and materials from which participants could freely choose according to their own preferences.

More than half of the studies (n=15; Multimedia Appendix 3) included synchronous or asynchronous personal contact with the facilitator via videoconferences, telephone calls, or emails to get regular feedback on their practice in the program.

The intervention duration was usually 8 weeks (n=10), with a session frequency of once per week, ranging from 2 weeks to 6 months. Of all 24 studies, only 2 included a retreat day [62,95].

Risk of Bias Assessment

The results of the assessment of the risk of bias in the included studies are reported in Multimedia Appendix 3. Of the 24 included studies, 3 did not report all outcomes [75,82,91], 9 had incomplete outcome data (<15% attrition) or did not use ITT analysis [54,71,75,76,78,80,82,85,92], and 10 did not have a comparison group (pre-post studies mentioned earlier). Of the 14 RCT studies, 2 were categorized as unclear regarding random sequence allocation and allocation concealment [85,92] and 2 were categorized as unclear regarding the similarity of groups at baseline or using an appropriate adjustment [87,89]. None of the studies met the criteria for blinding of participants, personnel, or outcome assessment, resulting in a high risk of bias (note that because of the nature of the included studies, the blinding of participants, personnel, and outcome assessment was not possible).

Feasibility

The reviewed studies examined feasibility in terms of attrition, retention, and adherence. All studies that examined feasibility as a primary outcome (n=12) concluded that eMBPs are feasible for patients with cancer [54,75,78-83,85,87,94,95]. In this review, we summarize the feasibility in terms of attrition and adherence.

Attrition

All 24 included studies reported clear data for computing the attrition rate. The attrition rate varied between 6% [79] and 46% [76,82], with an average of 25.3%. These in detail in Multimedia Appendix 3.

Predictors of Attrition

Regarding predictors of attrition, 9 studies reported the results of their analysis [54,62,71,73,76,77,80,89,91,93]. Participants who dropped out of the intervention (or did not respond to postassessment queries) had, in comparison with those who did not drop out, lower education [54,62,77], lower income [54], lower relationship satisfaction [76], lower baseline quality of life [77,93], higher baseline pain [93], shorter time since

diagnosis [54], brain metastasis [89], worse prognosis [77], less often breast cancer and more often *other* cancer types [77], and more often breast cancer and less often prostate cancer [91]; were more often younger [89], older [91], women [62,91], and men [77]; had comorbidities and were less often occupied by household activities [77]; and participated in an eHealth program rather than an in-person program [62]. However, the authors of 2 studies did not find any differences between completers and noncompleters [71,80]. In addition, the therapeutic alliance did not predict treatment dropout [73].

Adherence

There is vast heterogeneity in adherence measures in the reviewed studies. The studies have most often presented these measures of adherence as session completion (including session attendance), practice frequency (eg, how often participants meditated), and practice time (eg, time meditated). The summary is presented in Multimedia Appendix 3.

Session Completion

Regarding the session completion of all 24 included studies, only 14 studies could be assessed, 7 studies did not have predefined sessions [54,79,81,82,85,87,93], and 3 studies did not report session completion data [75,88,94].

Of these 14 studies, 6 had comparable results. The other 8 studies [71,76,78,80,83,89,91,92] are limited in terms of this comparison because of considerable heterogeneity in the type of reported participants (enrollers, completers, and unclear) and in the benchmark of minimum completed sessions (50%, 66%, and 100%). In the 6 comparable studies [62,77,84,86,90,95], at least half of the program was completed by 52% [86] to 83% [90] of the enrolled participants (on average across these studies, 70%).

There is a consensus among some authors [8,62] that the minimum adequate *dose attendance* of MBCT is 4 of 8 overall sessions (half of the program). In this context, based on the 6 comparable studies, eMBPs are on average feasible for 70% of participants.

Practice Frequency

Of all 24 included studies, 7 reported the practice frequency [54,78,80,82,88,93,94]. In only 2 studies [87,94], it was possible to derive the practice frequency adherence rate. In a study by Kubo et al [87], 50% of completers were adherent for at least 50%, and in a study by Russel et al [94], 61% to 80% of completers were 100% adherent. In the remaining 5 studies [62,78,82,88,93], the adherence rate could not be calculated because the minimum recommended time or the proportion of participants who were adherent was not defined.

Practice Time

Of all 24 included studies, 7 reported the practice time [62,78,81,85,91,93,95], but the practice time adherence rate could be derived only in 2 studies. In one study [81], 67% of completers were adherent to 100%, and in another study [85], 56% of completers were adherent to 70% of the recommended time.

In the remaining 5 studies [62,78,91,93,95], the adherence rate could not be calculated because the minimum recommended time or the proportion of participants who were adherent was not defined.

Predictors of Nonadherence

Regarding predictors of nonadherence, 5 studies reported the results of their analysis [76,77,82,84,91]. Participants who were nonadherent were more often men [76,77], were more often depressed at baseline [77], had less depressive symptoms [82], had a lower education [77], had a paid job less often [77], had no previous experience with mindfulness [77], and used sleeping medication less often [77]. In 2 studies [77,91], no differences in baseline characteristics were found between adherent and nonadherent participants.

Efficacy

The efficacy results of eMBPs in patients with cancer on all monitored variables are reported in Multimedia Appendix 3. The main effect of interventions on measured outcomes is presented in Figure 2. In the text of the results section, only the outcomes that were measured by at least four studies are reported; in studies with multiple follow-up endpoints, only the results at the postintervention and the last follow-up endpoint are reported. When multiple measurement instruments were used to measure the same outcome domain within the same time frame, the average effect estimate was calculated. Note that negative effect sizes are indicative of beneficial effects of interventions for outcomes: stress-related symptoms, anxiety and depression, fatigue, sleep problems, and pain.

Figure 2. The effect sizes of eHealth mindfulness-based programs on measured outcomes at postintervention.
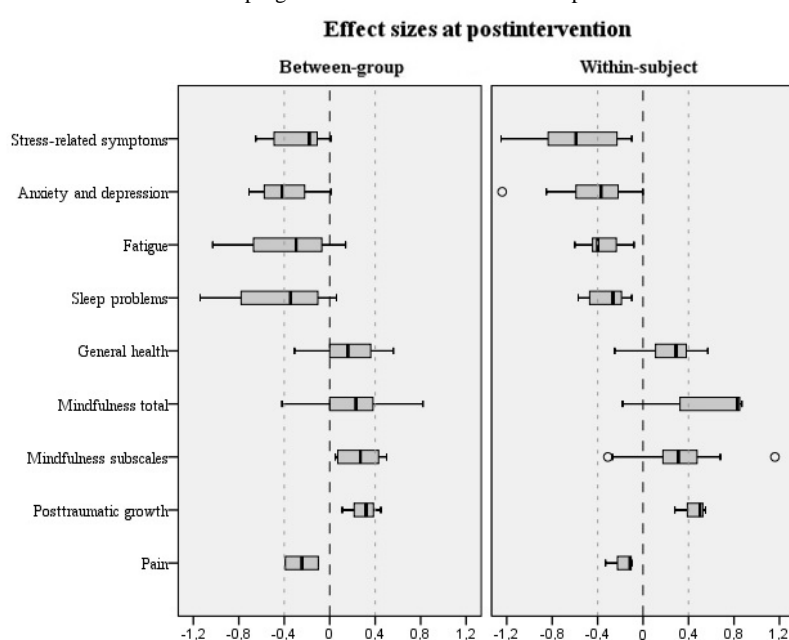
eMBPs were compared with active controls in 7 studies [62,84-86,89,90,92]. The Study Characteristics section provides the details of the type of active controls. Between-group results of eMBPs compared with no treatment and minimal treatment controls are summarized together, as differences consistently moderated by these control conditions by the phenomenological exploration were not found, and a recent review by Goldberg et al [17] showed that comparing MBPs against no treatment and minimal treatment controls did not differ; Cohen d<0.2. The results of comparisons to specific active interventions are reported below.

Stress-Related Symptoms

Results of Within-Subject Comparisons

The results at postintervention were reported by 8 studies [54,75,76,80,81,87,92,95]; Cohen d ranged from −1.25 to −0.10 (median −0.43, IQR −0.84 to −0.23). The results at follow-up were reported by only one study [76], where Cohen d of the female subgroup was −0.57 and that of the male subgroup was 0.01.

Results of Between-Group Comparisons

The results at postintervention were reported by 5 studies [71,87,91,94,95]; Cohen d ranged from −0.65 to 0.01 (median −0.18, IQR −0.49 to −0.11). The results at follow-up were reported by 4 studies [71,86,89,91]; Cohen d ranged from −0.68 to 0.05 (median −0.17, IQR −0.39 to −0.02).

Anxiety and Depression

Anxiety and depression outcomes are reported together for better summation, as some studies reported both anxiety and depression result in one merged total score [62,70,77,84].

Results of Within-Subject Comparisons

The results at postintervention were reported by 10 studies [54,75,76,78,80,81,83,85,87,92]; Cohen d ranged from −1.24 to −0.1 (median −0.38, IQR −0.62 to −0.27). The results at follow-up were reported by 2 studies [70,77], with Cohen d=−0.71 [77] and d=−0.32 [70].

Results of Between-Group Comparisons

The results at postintervention were reported by 4 studies [62,72,87,91]; Cohen d ranged from −0.71 to 0.01 (median

−0.42, IQR −0.58 to −0.22). The results at follow-up were reported by 5 studies [84,86,89-91]; Cohen d ranged from −0.53 to 0.14 (median −0.19, IQR −0.37 to −0.06).

Fatigue

Results of Within-Subject Comparisons

The results at postintervention were reported by 7 studies [54,75,81,83,85,87,92]; Cohen d ranged from −0.6 to −0.08 (median −0.40, IQR −0.43 to −0.25). The results at follow-up were reported by 1 study [77], with Cohen d=−1.45.

Results of Between-Group Comparisons

The results at postintervention were reported by 4 studies [71,85,87,88]; Cohen d ranged from −1.03 to 0.14 (median −0.30, IQR −0.49 to −0.18). The results at follow-up were reported by 3 studies [71,84,90]; Cohen d ranged from −0.69 to 0.07 (median −0.23, IQR −0.38 to −0.13).

Sleep Problems

Results of Within-Subject Comparisons

The results at postintervention were reported by 6 studies [54,75,80,81,85,87]; Cohen d ranged from −0.57 to −0.1 (median −0.27, IQR −0.44 to −0.19). None of the studies reported results at follow-up.

Results of Between-Group Comparisons

The results at postintervention were reported by 4 studies [72,87,88,91], Cohen d ranged from −1.14 to 0.06 (median −0.35, IQR −0.60 to −0.19). The results at follow-up were reported by 2 studies, with Cohen d=−0.32 [96] and d=0.04 [91].

General Health

General health includes all measurements assessing physical functioning, mental health, disability, and quality of life.

Results of Within-Subject Comparisons

The results at postintervention were reported by 6 studies [54,75,80,81,85,87]; Cohen d ranged from −0.25 to 0.57 (median 0.29, IQR 0.11 to 0.38). The results at follow-up were reported by only one study [62], with Cohen d=0.44.

Results of Between-Group Comparisons

The results at postintervention were reported by 6 studies [62,71,85,87,91,93]; Cohen d ranged from −0.31 to 0.56 (median 0.16, IQR 0.02 to 0.33). The results at follow-up were reported by 5 studies [71,86,90,91,93]; Cohen d ranged from −0.12 to 0.43 (median 0.10, IQR −0.10 to 0.22).

Mindfulness

Some studies reported the results of mindfulness measurements by total score, but others by subscales only (eg, observing, describing, acting with awareness, nonjudging, nonreacting). As these results are incomparable, they are reported separately. First, we report results by total scores, and then we report the results of the subscales only.

Results of Within-Subject Total Score Comparisons

The results at postintervention were reported by 3 studies [78,79,81], Cohen d ranged from −0.18 to 0.87 (median 0.33,

IQR 0.07 to 0.58). Neither of the studies reported results at follow-up.

Results of Between-Group Total Score Comparisons

The results at postintervention were reported in 5 studies [62,71,92-94]; Cohen d ranged from −0.42 to 0.82 (median 0.23, IQR 0.00 to 0.38). The results at follow-up were reported by 2 studies [71,93], with Cohen d=0.00 [71] and d=0.3 [93].

Results of Within-Subject Subscales Comparisons

The results at postintervention were reported by 4 studies [75,81,87,95]; Cohen d ranged from −0.31 to 1.16 (median 0.31, IQR 0.18 to 0.48). None of the studies reported results at follow-up.

Results of Between-Group Subscales Comparisons

The results at postintervention were reported by 2 studies [87,95]; Cohen d ranged from 0.05 to 0.5 (median 0.27, IQR 0.10 to 0.42). The results at follow-up were reported by only one study [86]; Cohen d ranged from 0.03 to 0.16.

Posttraumatic Growth

Results of Within-Subject Comparisons

The results at postintervention were reported by 3 studies [78,87,95]; Cohen d ranged from 0.28 to 0.55 (median 0.42, IQR 0.35 to 0.48). All 3 studies [78,87,95] reported statistically significant improvements. None of the studies reported results at follow-up.

Results of Between-Group Comparisons

The results at postintervention were reported by 3 studies [71,87,95]; Cohen d ranged from 0.11 to 0.45 (median 0.32, IQR 0.22-0.39). A statistically significant improvement (in favor of the intervention group) was reported by only one study [71]. The results at follow-up were reported by 2 studies [71,86], with Cohen d=0.15 [86] and d=0.21 [71]. A statistically significant difference (improvement in favor of the intervention group) was reported by 1 study [71].

Pain

Results of Within-Subject Comparisons

The results at postintervention were reported by 4 studies [81,83,85,87]; Cohen d ranged from −0.33 to −0.10 (median −0.12, IQR −0.17 to −0.11). Neither of the studies reported results at follow-up.

Results of Between-Group Comparisons

The results at postintervention were reported by 2 studies, with Cohen d=−0.39 [87] and d=−0.1 [85]. The results at follow-up were reported by 1 study [90], with Cohen d=−0.10. The reported difference was not statistically significant.

Adverse Effects

Some small adverse effects (Cohen d=0.2-0.49) were found in some studies on mindfulness (d=−0.42) [94] related to acting with awareness (d=−0.31), nonjudging (d=−0.27) [75], sex-related distress (d=0.26), depression (d=0.2) in males [76], satisfaction with sexual life (d=−0.22), mental health (d=−0.25) [85], and social support (d=−0.22) [79].

eMBP Versus Specific Active Interventions

eMBPs were compared with specific active interventions by 4 studies, published in 5 papers [62,70,84,89,92].

Compen et al [62] compared eMBCT with parallel, face-to-face MBCT and reported that eMBCT is superior for improving fear of cancer recurrence at postintervention ($d=-0.21$) and for improving anxiety and depression at follow-up ($d=-0.22$ [70]). For other outcomes, including rumination, quality of life, and mindfulness, they found comparable results ($d<0.2$) at both postintervention and follow-up time points.

Bruggeman-Everts et al [84] compared web-based eMBCT tailored for improving chronic cancer–related fatigue with physiotherapist-guided AAF therapy encompassing the use of an accelerometer. They reported that eMBCT is inferior for improving fatigue ($d=0.37$) at follow-up. For other outcomes, including positive and negative effects, anxiety, and depression, they found comparable results ($d<0.2$).

Price-Blackshear et al [92], who compared eMBSR completed only by patients with parallel eMBSR completed by couples (patients together with their partners) at postintervention, reported that eMBSR is inferior for improving anxiety ($d=-0.27$), depression ($d=0.38$), stress-related symptoms ($d=-0.38$), and mindfulness ($d=-0.63$), but surprisingly, eMBSR is superior for improving dyadic adjustment ($d=0.39$) and quality of marriage ($d=0.29$). They found comparable results for fatigue and interpersonal mindfulness ($d<0.2$).

Milbury et al [89] compared a couple-based meditation program with a supportive-expressive group at 3-month follow-up and reported that the meditation program was superior for improving depression ($d=0.59$) and cancer-related stress symptoms ($d=0.54$). They found comparable results for spiritual well-being ($d<0.2$).

Predictors, Moderators, Mediators, and Working Mechanisms of eMBPs Efficacy

Regarding moderators of eMBP efficacy, 9 studies reported the results of their analysis [62,72,73,75,76,80,87,88,91,97]. Participants who reported greater improvements in at least one of the reported outcomes (stress symptoms, depression and anxiety, spirituality, mindfulness skills, posttraumatic growth, quality of life, sleep problems, and pain) were younger [97], were more adherent [54,80,88], were male [97], were female [76], had prior exposure to meditation [75], participated in tandem with their caregiver [75], and reported higher baseline neuroticism [62], poorer baseline global mental health [72], and early therapeutic alliance [73]. No moderation effect was found for cancer stage [97], type [91], or, in contrast to other studies, adherence [91].

Regarding mediators of eMBP efficacy, 2 studies reported the results of their analysis [70,71]. Mindfulness skills [70,71], acceptance [71], fear of cancer recurrence [70], and rumination [70] were found.

Discussion

Principal Findings

This systematic review aims to examine the feasibility and efficacy of eMBPs in improving mental health and well-being in patients with cancer, to describe intervention characteristics and delivery modes of these programs, and to summarize the results of the included studies regarding the moderators, mediators, and predictors of efficacy, adherence, and attrition. Although vast heterogeneity in the intervention and population characteristics was found in the reviewed literature, most of the studies suggested that eHealth is an appropriate way to deliver proven mindfulness effects to patients with cancer. In general, the reviewed studies' results revealed that eMBPs have the potential to improve various outcomes. Some studies suggest that the effects may be maintained long-term after the end of the intervention. In addition, eMBP is equally effective (at postintervention) as a web-based behavior intervention [62,84] and even more effective (at long-term follow-up visits) as original face-to-face MBCT [70] for psychological distress.

Some significant predictors and moderators of attrition, adherence, and efficacy were found among both the participants and the applications across studies; however, any conclusion would be premature.

Although the results of this review are promising, the small number of RCT studies per outcome, substantial variability across studies, nondifferentiation between primary and secondary outcomes in results summarization, and lack of meta-analysis warrant caution in interpreting and generalizing the observed effects and relationships, as they could be overestimated. The results are heterogeneous across studies and vary between null and large effect sizes. This heterogeneity probably reflects the high variability in the population, intervention, and study characteristics. It also mirrors the fact that the bio-psychosocial complexity of oncological disease–induced distress is enormous, and patients differ in their needs. Existing eMBPs have mostly been trying to convert proven face-to-face mindfulness programs to eHealth mode. They have not yet exploited the full potential of eHealth technology options.

Attrition Rate

The attrition rate varied between 13% and 48% in the reviewed studies. This corresponds with the data presented in the systematic review of eMBPs in clinical and nonclinical populations, where the attrition rate varied between 7.7% and 52% [44]. The average attrition rate of face-to-face modes of mindfulness-based interventions ranges from 3% to 40% in clinical and nonclinical populations [98] or below 25% in most studies on adults with chronic medical diseases [99]. For comparison, in meaning-centered group psychotherapy for patients with advanced cancer in one study, 28.1% of patients dropped out before the start of the group and 24.5% of the participants dropped out after they began treatment [100]. It seems that the attrition rate of patients with cancer may be higher in eMBPs than in face-to-face programs. This is supported by the results by Compen et al [62], who documented that eHealth MBCT had a higher attrition rate than face-to-face MBCT.

XSL•FO
RenderX

Adherence

Measurement

The number of completed modules was the most prevalent measure of adherence. However, it is often unclear in most studies why certain measures of adherence were chosen over others, as the complete definition was missing in assessed studies. For comparison, Donkin et al [101] assembled a list of methods for measuring adherence to e-therapy (log-ins to the program, module completion, time spent on the internet, completion of a predefined activity or use of an internet tool, posts made, pages viewed, replies to emails, forum visits, self-reported completion of offline activities, and print requests made). In this context, the studies have not yet used available adherence markers, and it seems that many eMBP apps make little use of the current technology options. The majority of the abovementioned methods for measuring adherence to e-therapy can be measured automatically. Unfortunately, programming these app analytical tools is costly and time consuming.

If we use MBSR or MBCT programs as a gold standard in this intervention area, there is some consensus among experts, supported by some experimental data, that 4-week mindfulness programs seem to be efficacious for promoting well-being and stress reduction, and this amount of completion can be considered as a minimum adequate *dose* [8,62,102]. On average, two-thirds of the enrolled participants in the reviewed studies (52%-83%) completed at least half of the sessions. These results are similar to those from a review of web-based mindfulness interventions for people with physical health conditions [51], where the completion rate varied between 60% and 94%.

Efficacy

In general, the review confirmed the hypothesis that mindfulness delivered via eMBPs is able to induce a desirable change in subjectively assessed levels of stress, anxiety, depression, fatigue, sleep problems, mindfulness, posttraumatic growth, pain, and some parameters of general health. The direction of the found effects on most of the mentioned outcomes is consistent with recent meta-analysis results in face-to-face MBPs [23,70] and in eMBP [41].

The effect sizes were highly heterogeneous between studies, regardless of their methodological quality (Multimedia Appendix 3). The high heterogeneity in the efficacy of various outcomes is consistent with most of the literature related to eMBP [44,51], and it could be partially explained by high variability in the interventions, selected populations, and other study characteristics.

A small adverse effect was described in some studies. From a clinical perspective, the adverse effect on mindfulness [75,94], depression [76], some parameters of mental health [85], and social support [79] should be taken seriously. A recent systematic review of eMBPs in patients with medical conditions found no adverse effect on measured outcomes [60]. Another similar review did not report these data [51]. Although these findings are relatively marginal in our review, they raise a crucial question concerning the individuals for whom eMBP is appropriate and for whom it is not. This question is all the more important because there is usually no control over the mental state of patients, especially in anonymous eMBP, in comparison with face-to-face MBPs. Mindfulness training is paradoxical, and the instruction to focus directly on negative emotions goes against the inherent tendency to avoid unpleasant stimuli. To accept reality, whatever it is, represents the essence of the mindfulness approach. Mindfulness practice supports being in contact with whatever appears in an open, accepting, curious, and nonjudgmental manner. It is easy to imagine that some patients are not ready for this kind of mature emotion regulation strategy, such as acceptance and nonjudgmental openness to experience. After all, some patients experiencing extreme stress associated with cancer diagnosis and treatment involuntarily activate automatic psychological processes as defense mechanisms to reduce the anxiety of the painful emotions related to the illness [103,104]. Thus, mindfulness training can go against this self-protective strategy and, in some individuals, may worsen anxiety and depression and decrease mindfulness.

The selection of suitable patients is addressed in standard face-to-face MBPs in a personal interview before its start. However, there is still no clear consensus for whom participation in an MBP may be contraindicated [105]. In eMBPs, this is addressed in only some studies in the form of a phone call or questionnaire. The majority of studies relied solely on exclusion criteria presented in the input questionnaire at the start of the program. In the context of eMBPs, caution is in order.

A relatively small effect was observed in the context of pain relief, especially given that the pain is prevalent in 50.7% of patients with cancer [106] and that stress and pain reduction is the primary goal of MBSR [107] and depression in MBCT [108]. However, these results follow face-to-face mindfulness-based intervention efficacy. A recent meta‐analysis of RCT studies by Cillessen et al [109] also reported a small mindfulness effect on pain (Hedges g ranged between 0.18 and 0.20). More generally, Warth et al [110] reported a small effect of psychological intervention (relaxation, cognitive behavioral therapy, music therapy, mindfulness-based and acceptance-based interventions, and supportive-expressive group therapy) on pain intensity ($d=-0.29$; 95% CI -0.54 to -0.05). For mindfulness-based interventions, it was $d=0.14$. For comparison with a less severe diagnosis such as migraine [111], the effect size of MBSR on pain was medium to large (sensory component 0.79 and affective component $d=0.81$). In the context of eMBP, Ljotsson et al [112] used internet-delivered exposure and mindfulness-based therapy for irritable bowel syndrome. The effect size on pain was medium ($d=0.64$). The lower effect of eMBP on pain could be explained as a result of cancer progression over time in many patients in the program. In this context, disease progression should be monitored.

Population and Intervention Characteristics Increasing Efficacy

Although there is some knowledge of what the predictors increasing eMBPs adherence and efficacy are, this review shows that only 9 studies analyzed this. Some predictors were mentioned in the results of one study; some of them have been reported repeatedly. Higher adherence within the eMBP protocol and participation in tandem with a caregiver or partner were associated with greater improvements in measured outcomes.

Increased mindfulness was also a repeatedly reported mediator of the eMBP effect [70,71]. This finding is consistent with studies documenting mindfulness as one of the main mediators of its effects across the measured outcomes [113-115]. These results underline the importance of adherence. However, the analysis of adherence predictors was presented in only 5 reviewed studies.

Participants Characteristics

The 3 predictors of adherence that are mentioned most often in the eHealth research area are age, gender, and education [116-119]; however, findings seemed mixed in patients with cancer [120]. The reviewed studies support this trend with the prediction of gender by 2 studies [76,77] and education by 1 study [77]. This is consistent with findings from other eHealth interventions [121,122] and broader research on health behaviors, indicating that women are more likely to engage in such interventions than men [123]. For men, higher adherence was found in face-to-face psychotherapy [124], which suggests gender preference in different formats of psychological therapy. The role of age-adherence association was not reported in any of the reviewed studies.

The most comprehensive analysis of moderators and predictors from reviewed studies was by Cillessen et al [70]. The investigation revealed that lower levels of psychological distress, rumination, and neuroticism and a higher level of extraversion and agreeableness at the start of the eMBCT and MBCT program predicted less psychological distress at the 9-month follow-up after both interventions. The program-induced changes in mindfulness skills, fear of cancer recurrence, and rumination during both interventions predicted less psychological distress at follow-up. As Cillessen et al [70] discussed, most of these results are in accordance with previous research. For example, a study by Lengacher et al [125] found that patients with more baseline severity had more severe complaints at follow-up (however, patients with more baseline severity benefit relatively more from MBPs than patients with less severity). Other studies [113-115] identified fear of cancer recurrence, rumination, and mindfulness skills as mediators. However, although the authors discuss explanations of why personality traits predict MBP efficacy, some authors have documented the absence of moderation effects of personality traits. Mixed results in terms of the moderation role of personality traits were reported in a review by Vibe et al [126].

Intervention Characteristics

We agree with Keleders et al [127] that the critical question in this area concerns which characteristics of web-based or mobile app interventions related to technology and interaction are linked to better adherence.

We found considerable heterogeneity in the platforms and delivery modes of eMBPs across the studies, which makes it impossible to systematically report which of them are associated with the best efficacy. We share this conclusion with other authors [41,51]. In this context, the comparison of different delivery modes of eMBPs, including face-to-face programs in one arm, is still missing. The exception is the study by Compen

et al [62], in which the adherence of patients did not differ between eMBCT and MBCT.

A reminder system using email, text messages, or messages on a smartphone is a unique option of eHealth technology [128,129]. The utility of the reminder system also corresponds to our clinical experience with frequent patient statements of how great it would be if somebody could remind them of essential things from their psychotherapy sessions at the right moment in their daily life before problematic behaviors, thoughts, and emotions appeared. Surprisingly, reminders are not widespread across reviewed studies (Table 1), and the considerable heterogeneity of their type, frequency, and content does not allow any conclusion about the feasibility and efficacy of reminders. Wells et al [130] documented the importance and efficacy of smart messaging that reminded oncology patients in an MBCT program of prescribed between‐session activities. The odds of program completion were 8 times greater for patients using smart messaging than for nonusers. There has not yet been a study comparing intervention arms with and without reminders in the field of eMBPs. The utility of the reminder system is nevertheless apparent.

In the context of the beneficial intervention factors reported by Bruggeman-Everts et al [84] from patient feedback, the results of a qualitative study by Compen et al [131] show that the same elements and others (eg, own time management, individual and home setting, website delivery) were reported by patients as beneficial. On the other hand, when patients mentioned a certain aspect as facilitating (eg, the individual setting, not having to cope with other patients' stories), they also mentioned it as a barrier (no peer support). This was also the case for timing, the individual nature, the asynchronous nature, the diaries, and the importance of self-discipline [131]. As the authors suggest, this ambiguity emphasizes the importance of offering some flexibility in eMBPs [132], so each participant could choose the intervention features and delivery modes according to their own preferences.

Participation in tandem can partially replace group support and facilitation usually present in face-to-face MBPs. Face-to-face MBPs provide patients with social support from others in the group and the development of a therapeutic alliance with lector. The therapeutic alliance, a common factor in psychotherapy, is thought to be an essential factor in its outcomes [133], and it is crucial in mindfulness programs as well [134]. Bisseling et al [73] found that therapeutic alliances are a significant predictor of eMBP efficacy. Although, we could not make any exact conclusions, as research and discussion are still in their early stages [135]; the social support and therapeutic alliance in eMBPs are a topic of clinical relevance for developers and providers of eHealth interventions. The development of technology provides a new practical tool for clinicians and psychologists to be able to take care of patients who are not physically present. Patients can go through the program in the comfort and safety of their homes, at their right time in the context of their medical restrictions, in an anonymous mode (if they want to and if the program allows it), and they can experience it with the real feeling of some kind of alliance with their doctors and psychologists in the treatment journey. In this mode, the therapeutic alliance is not a dyadic but triadic

relationship among the user, the e–mental health program, and the program supporter [135]. Some data in the literature even indicate that a therapeutic alliance with an asynchronous e–mental health program can also be stated [136]. Appointments, homework reminders, assessment, and feedback may also help to develop and foster the therapeutic alliance [137]. eMBPs can reduce social isolation and feelings of loneliness, and participation in tandem with a caregiver or a partner is a promising eMBPs option. A synchronous web-based program, for example, the teleconference, allows patients to be in contact with a lector and with the other program participants. Asynchronous eMBPs can offer participants the opportunity to chat with other participants directly via the app or via social networks or some blogging platforms. In the context of adherence and development of a therapeutic alliance, our own clinical experience has led us to prefer a program that is organized and provided by the concrete hospital or center where patients are treated. It allows them to know who is behind the program and enables them to be in contact while giving them a sense of hospital safety. A combination of personal recommendations from their clinicians or clinical psychologists summarized in handouts and short videos presented in hospital waiting rooms with links to the social media platform and web page of the program can create a field where trust, adherence, and therapeutic alliance can emerge, thus increasing the efficacy of the eMBP programs.

Risk of Bias

This review included both RCT and pre-post studies. Half of the RCT studies (7/14, 50%) and nearly half of the pre-post studies (4/10, 40%) all have measured biases classified as low. In the pre-post studies, a higher risk of bias could be expected because of the predominant pilot designs.

In the context of RCT studies, high-quality studies are needed to establish the effectiveness of eMBPs. The selected criteria of a 15% attrition cutoff for risk of bias assessments in the current systematic review could seem quite conservative for eHealth studies. The mean attrition rate in the reviewed studies was 25%. The mitigation of this requirement would lower the risk of bias for 1 study [71].

Limitations

The results of this review have some limitations. First, despite the growing empirical literature on the efficacy of eMBPs in patients with cancer, only a relatively small number of RCT studies have been published, so this review has a relatively small number of RCT studies per outcome.

Second, a summarization of effect estimates was used as a synthesizing method rather than a meta-analysis. This method does not account for differences in the relative sizes of the studies, and the performance of statistics applied in the context of summarizing effect estimates has not been evaluated.

Third, the effectiveness of the included studies varied considerably in terms of outcome, which may be explained by variability in study characteristics, such as participants with different diagnoses and their staging, heterogeneous intervention types (eg, ACT, MBSR, MBCT), various modes of delivery, and outcome measures.

Finally, this review does not differentiate between primary and secondary outcomes to summarize the results.

The relatively small number of RCT studies per outcome, lack of meta-analysis, substantial variability across studies, and nondifferentiation between primary and secondary outcomes in result summarization warrants caution in interpreting and generalizing the observed effects and relationships.

Future Directions

We are still at the beginning of answering the questions for whom, with what kind of suffering, in what period of life and disease, and at what readiness to change unhealthy behavior and regulate emotions eMBP is the best tool. We have increasing evidence that eMBPs are useful. However, it must be noted that eMBPs are only one of many other tools in the repertoire of the experienced clinician or clinical psychologist. The effect of eMBP could be maximized when it is recommended to patients suitable for this kind of intervention and in the most effective delivery mode. To answer these questions, we need to provide eMBP to a large sample of patients in properly designed RCT studies in which to manipulate various variables in different study arms (eg, reminders, introductory lectures with the facilitator, internet chat, web, app, or combinations, rewards, etc) in different patient subgroups (eg, by stage of cancer, by readiness to change their behavior and attitudes, with or without psychotherapy). Future research should verify beneficial effects and their moderators found in this exploratory review within RCT studies or within a review incorporating standard meta-analytical methods when more suitable RCT studies will be published.

Conclusions

To our knowledge, this is the first systematic review that evaluates the feasibility and efficacy of eMBPs in patients with cancer on various psychological and somatic outcomes. The results show that eMBPs are feasible and may be effective in improving various outcomes, especially anxiety and depression and posttraumatic growth. Thus, eHealth represents an appropriate way for mindfulness programs to be delivered to patients with cancer, and they may be even more effective than standard group face-to-face MBPs in some cases. Regarding moderators, a preliminary phenomenological exploration showed possibly important population and intervention factors, such as age, gender, and delivery mode. Although the results of this review are promising, it is still necessary to be wary in interpreting and generalizing the observed effects and relationships.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 11 KB - jmir_v22i11e20709_app1.pdf]

Multimedia Appendix 2
Characteristics of included studies (Table 2).
[PDF File (Adobe PDF File), 122 KB - jmir_v22i11e20709_app2.pdf]

Multimedia Appendix 3
Results of included studies (Table 3) and Risk of bias assessment (Table 4).
[PDF File (Adobe PDF File), 236 KB - jmir_v22i11e20709_app3.pdf]

References

1.   Cancer. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/cancer [accessed 2020-04-23]
2.   Sellick SM, Edwardson AD. Screening new cancer patients for psychological distress using the hospital anxiety and depression scale. Psychooncology 2007 Jun;16(6):534-542. [doi: 10.1002/pon.1085] [Medline: 16989004]
3.   Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. J Natl Cancer Inst 2009 Nov 4;101(21):1464-1488 [FREE Full text] [doi: 10.1093/jnci/djp336] [Medline: 19826136]
4.   Grassi L, Sabato S, Rossi E, Biancosino B, Marmai L. Use of the diagnostic criteria for psychosomatic research in oncology. Psychother Psychosom 2005;74(2):100-107. [doi: 10.1159/000083168] [Medline: 15741759]
5.   Carlson LE, Tamagawa R, Stephen J, Doll R, Faris P, Dirkse D, et al. Tailoring mind-body therapies to individual needs: patients' program preference and psychological traits as moderators of the effects of mindfulness-based cancer recovery and supportive-expressive therapy in distressed breast cancer survivors. J Natl Cancer Inst Monogr 2014 Nov;2014(50):308-314. [doi: 10.1093/jncimonographs/lgu034] [Medline: 25749597]
6.   Světláková L, Sláma O, Světlák M, Pochop L, Šedo J, Alexandrová R, et al. Prevalence of anxiety and depression and their impact on the quality of life of cancer patients treated with palliative antineoplasic therapy - results of the palint trial. Klin Onkol 2019;32(3):201-207. [doi: 10.14735/amko2019201] [Medline: 31216852]
7.   Carlson LE, Angen M, Cullum J, Goodey E, Koopmans J, Lamont L, et al. High levels of untreated distress and fatigue in cancer patients. Br J Cancer 2004 Jun 14;90(12):2297-2304 [FREE Full text] [doi: 10.1038/sj.bjc.6601887] [Medline: 15162149]
8.   Crane RS, Brewer J, Feldman C, Kabat-Zinn J, Santorelli S, Williams JM, et al. What defines mindfulness-based programs? The warp and the weft. Psychol Med 2017 Apr;47(6):990-999. [doi: 10.1017/S0033291716003317] [Medline: 28031068]
9.   Waller A, Williams A, Groff SL, Bultz BD, Carlson LE. Screening for distress, the sixth vital sign: examining self-referral in people with cancer over a one-year period. Psychooncology 2013 Feb;22(2):388-395. [doi: 10.1002/pon.2102] [Medline: 22135205]
10. Söllner W, DeVries A, Steixner E, Lukas P, Sprinzl G, Rumpold G, et al. How successful are oncologists in identifying patient distress, perceived social support, and need for psychosocial counselling? Br J Cancer 2001 Jan;84(2):179-185 [FREE Full text] [doi: 10.1054/bjoc.2000.1545] [Medline: 11161373]
11. Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. Gen Hosp Psychiatry 1982 Apr;4(1):33-47. [doi: 10.1016/0163-8343(82)90026-3] [Medline: 7042457]

12. Segal Z, Williams J, Teasdale J. Mindfulness-based Cognitive Therapy for Depression: a New Approach to Preventing Relapse. New York, USA: Guilford Press; 2002.

13. Baer R. Mindfulness-Based Treatment Approaches: Clinician's Guide to Evidence Base and Applications Internet. London, UK: Academic Press; 2014.

14. Bishop SR. Mindfulness: a proposed operational definition. Clin Psychol: Sci Pract 2004 Aug 1;11(3):230-241. [doi: 10.1093/clipsy/bph077]

15. Keng S, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]

16. Tomlinson ER, Yousaf O, Vittersø AD, Jones L. Dispositional mindfulness and psychological health: a systematic review. Mindfulness (N Y) 2018;9(1):23-43 [FREE Full text] [doi: 10.1007/s12671-017-0762-6] [Medline: 29387263]

17. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: a systematic review and meta-analysis. Clin Psychol Rev 2018 Feb;59:52-60 [FREE Full text] [doi: 10.1016/j.cpr.2017.10.011] [Medline: 29126747]

18. Rahimi-Ardabili H, Reynolds R, Vartanian LR, McLeod LV, Zwar N. A systematic review of the efficacy of interventions that aim to increase self-compassion on nutrition habits, eating behaviours, body weight and body image. Mindfulness 2017 Sep 20;9(2):388-400. [doi: 10.1007/s12671-017-0804-0]

19. Baxter J, Welsh H, Grayer J. Mindfulness-based interventions for cancer-related pain and depression: a narrative review of current evidence and future potential. Curr Opin Support Palliat Care 2019 Jun;13(2):81-87. [doi: 10.1097/SPC.0000000000000428] [Medline: 30973395]

20. Shennan C, Payne S, Fenlon D. What is the evidence for the use of mindfulness-based interventions in cancer care? A review. Psychooncology 2011 Jul;20(7):681-697. [doi: 10.1002/pon.1819] [Medline: 20690112]

21. Ott MJ, Norris RL, Bauer-Wu SM. Mindfulness meditation for oncology patients: a discussion and critical review. Integr Cancer Ther 2006 Jun;5(2):98-108. [doi: 10.1177/1534735406288083] [Medline: 16685074]

22. Smith JE, Richardson J, Hoffman C, Pilkington K. Mindfulness-Based Stress Reduction as supportive therapy in cancer care: systematic review. J Adv Nurs 2005 Nov;52(3):315-327. [doi: 10.1111/j.1365-2648.2005.03592.x] [Medline: 16194185]

23. Piet J, Würtzen H, Zachariae R. The effect of mindfulness-based therapy on symptoms of anxiety and depression in adult cancer patients and survivors: a systematic review and meta-analysis. J Consult Clin Psychol 2012 Dec;80(6):1007-1020. [doi: 10.1037/a0028329] [Medline: 22563637]

24. Cramer H, Lauche R, Paul A, Dobos G. Mindfulness-based stress reduction for breast cancer-a systematic review and meta-analysis. Curr Oncol 2012 Oct;19(5):e343-e352 [FREE Full text] [doi: 10.3747/co.19.1016] [Medline: 23144582]

25. Zainal NZ, Booth S, Huppert FA. The efficacy of mindfulness-based stress reduction on mental health of breast cancer patients: a meta-analysis. Psychooncology 2013 Jul;22(7):1457-1465. [doi: 10.1002/pon.3171] [Medline: 22961994]

26. Ledesma D, Kumano H. Mindfulness-based stress reduction and cancer: a meta-analysis. Psychooncology 2009 Jun;18(6):571-579. [doi: 10.1002/pon.1400] [Medline: 19023879]

27. Xunlin NG, Lau Y, Klainin-Yobas P. The effectiveness of mindfulness-based interventions among cancer patients and survivors: a systematic review and meta-analysis. Support Care Cancer 2020 Apr;28(4):1563-1578. [doi: 10.1007/s00520-019-05219-9] [Medline: 31834518]

28. Haller H, Winkler MM, Klose P, Dobos G, Kümmel S, Cramer H. Mindfulness-based interventions for women with breast cancer: an updated systematic review and meta-analysis. Acta Oncol 2017 Dec;56(12):1665-1676. [doi: 10.1080/0284186X.2017.1342862] [Medline: 28686520]

29. Ngamkham S, Holden JE, Smith EL. A systematic review: mindfulness intervention for cancer-related pain. Asia Pac J Oncol Nurs 2019;6(2):161-169 [FREE Full text] [doi: 10.4103/apjon.apjon_67_18] [Medline: 30931361]

30. Zimmermann FF, Burrell B, Jordan J. The acceptability and potential benefits of mindfulness-based interventions in improving psychological well-being for adults with advanced cancer: A systematic review. Complement Ther Clin Pract 2018 Feb;30:68-78. [doi: 10.1016/j.ctcp.2017.12.014] [Medline: 29389483]

31. Coutiño-Escamilla LPTFDSIBCPSRAMAORCT, Piña-Pozas M, Tobías Garces A, Gamboa-Loira B, López-Carrillo L. Non-pharmacological therapies for depressive symptoms in breast cancer patients: Systematic review and meta-analysis of randomized clinical trials. Breast 2019 Apr;44:135-143. [doi: 10.1016/j.breast.2019.01.006] [Medline: 30776733]

32. Hall DL, Luberto CM, Philpotts LL, Song R, Park ER, Yeh GY. Mind-body interventions for fear of cancer recurrence: a systematic review and meta-analysis. Psychooncology 2018 Nov;27(11):2546-2558 [FREE Full text] [doi: 10.1002/pon.4757] [Medline: 29744965]

33. Carlson LE, Beattie TL, Giese-Davis J, Faris P, Tamagawa R, Fick LJ, et al. Mindfulness-based cancer recovery and supportive-expressive therapy maintain telomere length relative to controls in distressed breast cancer survivors. Cancer 2015 Feb 1;121(3):476-484 [FREE Full text] [doi: 10.1002/cncr.29063] [Medline: 25367403]

34. Bartley T. A Kindly Approach to Being with Cancer. New York, USA: John Wiley & Sons; 2016.

35. Carlson L, Speca M. Mindfulness-based cancer recovery: a step-by-step MBSR approach to help you cope with treatment & reclaim your life. Oakland, CA: New Harbinger Publications; 2011.

XSL•FO

RenderX

36. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]

37. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

38. Prochaska J, Norcross J. Systems of Psychotherapy: A Transtheoretical Analysis. 5th Edition. Pacific Grove, CA: Brooks/Cole Pub; 2003.

39. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]

40. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]

41. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]

42. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

43. Cavanagh K, Churchard A, O'Hanlon P, Mundy T, Votolato P, Jones F, et al. A randomised controlled trial of a brief online mindfulness-based intervention in a non-clinical population: replication and extension. Mindfulness (N Y) 2018;9(4):1191-1205 [FREE Full text] [doi: 10.1007/s12671-017-0856-1] [Medline: 30100934]

44. Fish J, Brimson J, Lynch S. Mindfulness interventions delivered by technology without facilitator involvement: what research exists and what are the clinical outcomes? Mindfulness (N Y) 2016;7(5):1011-1023 [FREE Full text] [doi: 10.1007/s12671-016-0548-2] [Medline: 27642370]

45. Querstret D, Cropley M, Fife-Schaw C. The effects of an online mindfulness intervention on perceived stress, depression and anxiety in a non-clinical sample: a randomised waitlist control trial. Mindfulness (N Y) 2018;9(6):1825-1836 [FREE Full text] [doi: 10.1007/s12671-018-0925-0] [Medline: 30524514]

46. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015 Nov;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]

47. Krusche A, Cyhlarova E, Williams JM. Mindfulness online: an evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open 2013 Nov 29;3(11):e003498. [doi: 10.1136/bmjopen-2013-003498] [Medline: 24293203]

48. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]

49. Dowd H, Hogan MJ, McGuire BE, Davis MC, Sarma KM, Fish RA, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation: a randomized controlled study. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]

50. Davis MC, Zautra AJ. An online mindfulness intervention targeting socioemotional regulation in fibromyalgia: results of a randomized controlled trial. Ann Behav Med 2013 Dec;46(3):273-284. [doi: 10.1007/s12160-013-9513-7] [Medline: 23670111]

51. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]

52. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]

53. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Dirkse DA, et al. The eCALM Trial-eTherapy for cancer appLying mindfulness: online mindfulness-based cancer recovery program for underserved individuals living with cancer in Alberta: protocol development for a randomized wait-list controlled clinical trial. BMC Complement Altern Med 2013 Feb 16;13:34 [FREE Full text] [doi: 10.1186/1472-6882-13-34] [Medline: 23414206]

54. Kubo A, Altschuler A, Kurtovich E, Hendlish S, Laurent CA, Kolevska T, et al. A pilot mobile-based mindfulness intervention for cancer patients and their informal caregivers. Mindfulness (N Y) 2018 Dec;9(6):1885-1894 [FREE Full text] [doi: 10.1007/s12671-018-0931-2] [Medline: 30740187]

55. Bartley T. A Kindly Approach to Being with Cancer. New York, USA: John Wiley & Sons; 2016.

56. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]

57. Linehan M. Cognitive-Behavioral Treatment of Borderline Personality Disorder. New York, USA: Guilford Publications; 2018.

58. Graham CD, Gouick J, Krahé C, Gillanders D. A systematic review of the use of acceptance and commitment therapy (ACT) in chronic disease and long-term conditions. Clin Psychol Rev 2016 Jun;46:46-58. [doi: 10.1016/j.cpr.2016.04.009] [Medline: 27176925]

59. González-Fernández S, Fernández-Rodríguez C. Acceptance and commitment therapy in cancer: review of applications and findings. Behav Med 2019;45(3):255-269. [doi: 10.1080/08964289.2018.1452713] [Medline: 29558259]

60. Mikolasek M, Berg J, Witt CM, Barth J. Effectiveness of mindfulness- and relaxation-based ehealth interventions for patients with medical conditions: a systematic review and synthesis. Int J Behav Med 2018 Feb;25(1):1-16. [doi: 10.1007/s12529-017-9679-7] [Medline: 28752414]

61. Crane C, Williams JM. Factors associated with attrition from mindfulness-based cognitive therapy in patients with a history of suicidal depression. Mindfulness (N Y) 2010 Mar;1(1):10-20 [FREE Full text] [doi: 10.1007/s12671-010-0003-8] [Medline: 21125023]

62. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L, van der Lee M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. J Clin Oncol 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/JCO.2017.76.5669] [Medline: 29953304]

63. Millegan J, Manschot B, Dispenzieri M, Marks B, Edwards A, Raulston V, et al. Leveraging iPads to introduce meditation and reduce distress among cancer patients undergoing chemotherapy: a promising approach. Support Care Cancer 2015 Dec;23(12):3393-3394. [doi: 10.1007/s00520-015-2891-6] [Medline: 26260727]

64. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]

65. Higgins J, Green S. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2017-08-23] [WebCite Cache ID 6svURitRL]

66. Higgins J, Altman D, Sterne J. Assessing Risk of Bias in Included Studies. Cochrane Handbook for Systematic Reviews of Interventions. 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2017-08-23] [WebCite Cache ID 6svURitRL]

67. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. New York, USA: Routledge; 1988.

68. Dimidjian S, Segal ZV. Prospects for a clinical science of mindfulness-based intervention. Am Psychol 2015 Oct;70(7):593-620 [FREE Full text] [doi: 10.1037/a0039589] [Medline: 26436311]

69. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. Clin Psychol Sci 2014 Jan 1;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]

70. Cillessen L, Schellekens MP, van de Ven MO, Donders AR, Compen FR, Bisseling EM, et al. Consolidation and prediction of long-term treatment effect of group and online mindfulness-based cognitive therapy for distressed cancer patients. Acta Oncol 2018 Oct;57(10):1293-1302. [doi: 10.1080/0284186X.2018.1479071] [Medline: 29932784]

71. Hawkes AL, Pakenham KI, Chambers SK, Patrao TA, Courneya KS. Effects of a multiple health behavior change intervention for colorectal cancer survivors on psychosocial outcomes and quality of life: a randomized controlled trial. Ann Behav Med 2014 Dec;48(3):359-370. [doi: 10.1007/s12160-014-9610-2] [Medline: 24722960]

72. Puzia ME, Huberty J, Eckert R, Larkey L, Mesa R. Associations between global mental health and response to an app-based meditation intervention in myeloproliferative neoplasm patients. Integr Cancer Ther 2020;19:1534735420927780. [doi: 10.1177/1534735420927780] [Medline: 32564631]

73. Bisseling E, Cillessen L, Spinhoven P, Schellekens M, Compen F, van der Lee M, et al. Development of the therapeutic alliance and its association with internet-based mindfulness-based cognitive therapy for distressed cancer patients: secondary analysis of a multicenter randomized controlled trial. J Med Internet Res 2019 Oct 18;21(10):e14065 [FREE Full text] [doi: 10.2196/14065] [Medline: 31628791]

74. Hawkes AL, Chambers SK, Pakenham KI, Patrao TA, Baade PD, Lynch BM, et al. Effects of a telephone-delivered multiple health behavior change intervention (CanChange) on health and behavioral outcomes in survivors of colorectal cancer: a randomized controlled trial. J Clin Oncol 2013 Jun 20;31(18):2313-2321. [doi: 10.1200/JCO.2012.45.5873] [Medline: 23690410]

75. Atreya CE, Kubo A, Borno HT, Rosenthal B, Campanella M, Rettger JP, et al. Being Present: a single-arm feasibility study of audio-based mindfulness meditation for colorectal cancer patients and caregivers. PLoS One 2018;13(7):e0199423 [FREE Full text] [doi: 10.1371/journal.pone.0199423] [Medline: 30036361]

76. Brotto LA, Dunkley CR, Breckon E, Carter J, Brown C, Daniluk J, et al. Integrating quantitative and qualitative methods to evaluate an online psychoeducational program for sexual difficulties in colorectal and gynecologic cancer survivors. J Sex Marital Ther 2017 Oct 03;43(7):645-662. [doi: 10.1080/0092623X.2016.1230805] [Medline: 27592509]

77. Bruggeman Everts FZ, van der Lee ML, de Jager Meezenbroek E. Web-based individual mindfulness-based cognitive therapy for cancer-related fatigue — a pilot study. Internet Interv 2015 May;2(2):200-213. [doi: 10.1016/j.invent.2015.03.004]

78. Campo RA, Bluth K, Santacroce SJ, Knapik S, Tan J, Gold S, et al. A mindful self-compassion videoconference intervention for nationally recruited posttreatment young adult cancer survivors: feasibility, acceptability, and psychosocial outcomes. Support Care Cancer 2017 Jun;25(6):1759-1768. [doi: 10.1007/s00520-017-3586-y] [Medline: 28105523]

79.   Donovan E, Martin SR, Seidman LC, Zeltzer LK, Cousineau TM, Payne LA, et al. A mobile-based mindfulness and social support program for adolescents and young adults with sarcoma: development and pilot testing. JMIR Mhealth Uhealth 2019 Mar 18;7(3):e10921 [FREE Full text] [doi: 10.2196/10921] [Medline: 30882352]

80.   Kinner EM, Armer JS, McGregor BA, Duffecy J, Leighton S, Corden ME, et al. Internet-based group intervention for ovarian cancer survivors: feasibility and preliminary results. JMIR Cancer 2018 Jan 15;4(1):e1 [FREE Full text] [doi: 10.2196/cancer.8430] [Medline: 29335233]

81.   Lengacher CA, Reich RR, Ramesar S, Alinat CB, Moscoso M, Cousin L, et al. Feasibility of the mobile mindfulness-based stress reduction for breast cancer (mMBSR(BC)) program for symptom improvement among breast cancer survivors. Psychooncology 2018 Feb;27(2):524-531. [doi: 10.1002/pon.4491] [Medline: 28665541]

82.   Mikolasek M, Witt CM, Barth J. Adherence to a mindfulness and relaxation self-care app for cancer patients: mixed-methods feasibility study. JMIR Mhealth Uhealth 2018 Dec 6;6(12):e11271 [FREE Full text] [doi: 10.2196/11271] [Medline: 30522990]

83.   Plumb Vilardaga JC, Winger JG, Teo I, Owen L, Sutton LM, Keefe FJ, et al. Coping skills training and acceptance and commitment therapy for symptom management: feasibility and acceptability of a brief telephone-delivered protocol for patients with advanced cancer. J Pain Symptom Manage 2020 Feb;59(2):270-278. [doi: 10.1016/j.jpainsymman.2019.09.005] [Medline: 31539599]

84.   Bruggeman-Everts FZ, Wolvers MD, van de Schoot R, Vollenbroek-Hutten MM, Van der Lee ML. Effectiveness of two web-based interventions for chronic cancer-related fatigue compared to an active control condition: results of the "'fitter na kanker' randomized controlled trial. J Med Internet Res 2017 Oct 19;19(10):e336 [FREE Full text] [doi: 10.2196/jmir.7180] [Medline: 29051138]

85.   Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]

86.   Chambers SK, Occhipinti S, Foley E, Clutton S, Legg M, Berry M, et al. Mindfulness-based cognitive therapy in advanced prostate cancer: a randomized controlled trial. J Clin Oncol 2017 Jan 20;35(3):291-297. [doi: 10.1200/JCO.2016.68.8788] [Medline: 27870567]

87.   Kubo A, Kurtovich E, McGinnis M, Aghaee S, Altschuler A, Quesenberry C, et al. A randomized controlled trial of mhealth mindfulness intervention for cancer patients and informal cancer caregivers: a feasibility study within an integrated health care delivery system. Integr Cancer Ther 2019;18:1534735419850634. [doi: 10.1177/1534735419850634] [Medline: 31092044]

88.   Messer D, Horan JJ, Larkey LK, Shanholtz CE. Effects of internet training in mindfulness meditation on variables related to cancer recovery. Mindfulness 2019 Jun 4;10(10):2143-2151. [doi: 10.1007/s12671-019-01182-y]

89.   Milbury K, Li Y, Durrani S, Liao Z, Tsao AS, Carmack C, et al. A mindfulness-based intervention as a supportive care strategy for patients with metastatic non-small cell lung cancer and their spouses: results of a three-arm pilot randomized controlled trial. Oncologist 2020 Jul 4:- epub ahead of print [FREE Full text] [doi: 10.1634/theoncologist.2020-0125] [Medline: 32621630]

90.   Mosher CE, Secinti E, Li R, Hirsh AT, Bricker J, Miller KD, et al. Acceptance and commitment therapy for symptom interference in metastatic breast cancer patients: a pilot randomized trial. Support Care Cancer 2018 Jun;26(6):1993-2004 [FREE Full text] [doi: 10.1007/s00520-018-4045-0] [Medline: 29327085]

91.   Nissen ER, O'Connor M, Kaldo V, Højris I, Borre M, Zachariae R, et al. Internet-delivered mindfulness-based cognitive therapy for anxiety and depression in cancer survivors: A randomized controlled trial. Psychooncology 2020 Jan;29(1):68-75 [FREE Full text] [doi: 10.1002/pon.5237] [Medline: 31600414]

92.   Price-Blackshear MA, Pratscher SD, Oyler DL, Armer JM, Cheng A, Cheng MX, et al. Online couples mindfulness-based intervention for young breast cancer survivors and their partners: A randomized-control trial. J Psychosoc Oncol 2020;38(5):592-611. [doi: 10.1080/07347332.2020.1778150] [Medline: 32552446]

93.   Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: a randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]

94.   Russell L, Ugalde A, Orellana L, Milne D, Krishnasamy M, Chambers R, et al. A pilot randomised controlled trial of an online mindfulness-based program for people diagnosed with melanoma. Support Care Cancer 2019 Jul;27(7):2735-2746. [doi: 10.1007/s00520-018-4574-6] [Medline: 30506103]

95.   Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]

96.   Mosher CE, Secinti E, Hirsh AT, Hanna N, Einhorn LH, Jalal SI, et al. Acceptance and commitment therapy for symptom interference in advanced lung cancer and caregiver distress: a pilot randomized trial. J Pain Symptom Manage 2019 Oct;58(4):632-644. [doi: 10.1016/j.jpainsymman.2019.06.021] [Medline: 31255586]

97.   Zernicke KA, Campbell TS, Speca M, Ruff KM, Flowers S, Tamagawa R, et al. The ecalm trial: etherapy for cancer applying mindfulness. Exploratory analyses of the associations between online mindfulness-based cancer recovery participation and

changes in mood, stress symptoms, mindfulness, posttraumatic growth, and spirituality. Mindfulness 2016 May 14;7(5):1071-1081. [doi: 10.1007/s12671-016-0545-5]

98. Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clin Psychol: Sci Pract 2006 May 11;10(2):125-143. [doi: 10.1093/clipsy.bpg015]

99. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: a meta-analysis. J Psychosom Res 2010 Jun;68(6):539-544. [doi: 10.1016/j.jpsychores.2009.10.005] [Medline: 20488270]

100. Applebaum AJ, Lichtenthal WG, Pessin HA, Radomski JN, Simay Gökbayrak N, Katz AM, et al. Factors associated with attrition from a randomized controlled trial of meaning-centered group psychotherapy for patients with advanced cancer. Psychooncology 2012 Nov;21(11):1195-1204 [FREE Full text] [doi: 10.1002/pon.2013] [Medline: 21751295]

101. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

102. Demarzo M, Montero-Marin J, Puebla-Guedea M, Navarro-Gil M, Herrera-Mercadal P, Moreno-González S, et al. Efficacy of 8- and 4-session mindfulness-based interventions in a non-clinical population: a controlled study. Front Psychol 2017;8:1343 [FREE Full text] [doi: 10.3389/fpsyg.2017.01343] [Medline: 28848465]

103. Zimmerman TN, Porcerelli JH, Arterbery VE. Defensive functioning in cancer patients, cancer survivors, and controls. Psychoanalytic Psychol 2019 Jul;36(3):259-262. [doi: 10.1037/pap0000225]

104. di Giuseppe M, Ciacchini R, Micheloni T, Bertolucci I, Marchi L, Conversano C. Defense mechanisms in cancer patients: a systematic review. J Psychosom Res 2018 Dec;115:76-86. [doi: 10.1016/j.jpsychores.2018.10.016] [Medline: 30470322]

105. Dobkin PL, Irving JA, Amar S. For whom may participation in a mindfulness-based stress reduction program be contraindicated? Mindfulness 2011 Oct 6;3(1):44-50. [doi: 10.1007/s12671-011-0079-9]

106. van den Beuken-van Everdingen MH, Hochstenbach LM, Joosten EA, Tjan-Heijnen VC, Janssen DJ. Update on prevalence of pain in patients with cancer: systematic review and meta-analysis. J Pain Symptom Manage 2016 Jun;51(6):1070-90.e9 [FREE Full text] [doi: 10.1016/j.jpainsymman.2015.12.340] [Medline: 27112310]

107. Kabat-Zinn J. Full Catastrophe Living How to Cope With Stress, Pain and Illness Using Mindfulness Meditation. London, UK: Piatkus; 2011.

108. Teasdale JD, Segal ZV, Williams JM, Ridgeway VA, Soulsby JM, Lau MA. Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. J Consult Clin Psychol 2000 Aug;68(4):615-623. [doi: 10.1037//0022-006x.68.4.615] [Medline: 10965637]

109. Cillessen L, Johannsen M, Speckens AE, Zachariae R. Mindfulness-based interventions for psychological and physical health outcomes in cancer patients and survivors: A systematic review and meta-analysis of randomized controlled trials. Psychooncology 2019 Dec;28(12):2257-2269 [FREE Full text] [doi: 10.1002/pon.5214] [Medline: 31464026]

110. Warth M, Zöller J, Köhler F, Aguilar-Raab C, Kessler J, Ditzen B. Psychosocial interventions for pain management in advanced cancer patients: a systematic review and meta-analysis. Curr Oncol Rep 2020 Jan 21;22(1):3. [doi: 10.1007/s11912-020-0870-7] [Medline: 31965361]

111. Simshäuser K, Lüking M, Kaube H, Schultz C, Schmidt S. Is mindfulness-based stress reduction a promising and feasible intervention for patients suffering from migraine? A randomized controlled pilot trial. Complement Med Res 2020;27(1):19-30. [doi: 10.1159/000501425] [Medline: 31390617]

112. Ljótsson B, Falk L, Vesterlund AW, Hedman E, Lindfors P, Rück C, et al. Internet-delivered exposure and mindfulness based therapy for irritable bowel syndrome--a randomized controlled trial. Behav Res Ther 2010 Jun;48(6):531-539. [doi: 10.1016/j.brat.2010.03.003] [Medline: 20362976]

113. Lengacher CA, Shelton MM, Reich RR, Barta MK, Johnson-Mallard V, Moscoso MS, et al. Mindfulness based stress reduction (MBSR(BC)) in breast cancer: evaluating fear of recurrence (FOR) as a mediator of psychological and physical symptoms in a randomized control trial (RCT). J Behav Med 2014 Apr;37(2):185-195 [FREE Full text] [doi: 10.1007/s10865-012-9473-6] [Medline: 23184061]

114. Labelle LE, Campbell TS, Faris P, Carlson LE. Mediators of mindfulness-based stress reduction (MBSR): assessing the timing and sequence of change in cancer patients. J Clin Psychol 2015 Jan;71(1):21-40. [doi: 10.1002/jclp.22117] [Medline: 25099479]

115. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]

116. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

117. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]

118. Montero-Marin J, Gaete J, Araya R, Demarzo M, Manzanera R, Álvarez de Mon M, et al. Impact of a blended web-based mindfulness programme for general practitioners: a pilot study. Mindfulness 2017 Jun 15;9(1):129-139. [doi: 10.1007/s12671-017-0752-8]

119. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]

120. Carlson LE, Ursuliak Z, Goodey E, Angen M, Speca M. The effects of a mindfulness meditation-based stress reduction program on mood and symptoms of stress in cancer outpatients: 6-month follow-up. Support Care Cancer 2001 Mar;9(2):112-123. [doi: 10.1007/s005200000206] [Medline: 11305069]

121. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]

122. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in internet-based interventions. Patient Prefer Adherence 2008 Feb 2;2:57-65 [FREE Full text] [Medline: 19920945]

123. Berrigan D, Dodd K, Troiano RP, Krebs-Smith SM, Barbash RB. Patterns of health behavior in US adults. Prev Med 2003 May;36(5):615-623. [doi: 10.1016/s0091-7435(02)00067-1] [Medline: 12689807]

124. Strauss JL, Guerra VS, Marx CE, Eggleston AM, Calhoun PS. Adherence and psychotherapy. In: Bosworth H, editor. Improving Patient Treatment Adherence. New York, USA: Springer; 2010.

125. Lengacher CA, Reich RR, Paterson CL, Ramesar S, Park JY, Alinat C, et al. Examination of Broad Symptom Improvement Resulting From Mindfulness-Based Stress Reduction in Breast Cancer Survivors: A Randomized Controlled Trial. J Clin Oncol 2016 Aug 20;34(24):2827-2834 [FREE Full text] [doi: 10.1200/JCO.2015.65.7874] [Medline: 27247219]

126. Vibe MD, Bjørndal A, Fattah S, Dyrdal GM, Halland E, Tanner‐Smith EE. Mindfulness‐based stress reduction (MBSR) for improving health, quality of life and social functioning in adults: a systematic review and meta‐analysis. Mindfulness‐based stress reduction (MBSR) for improving health, quality of life and social functioning in adults: a systematic review and meta‐analysis 2017 Nov 1. [doi: https://doi.org/10.4073/csr.2017.11]

127. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

128. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

129. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

130. Wells C, Malins S, Clarke S, Skorodzien I, Biswas S, Sweeney T, et al. Using smart-messaging to enhance mindfulness-based cognitive therapy for cancer patients: A mixed methods proof of concept evaluation. Psychooncology 2020 Jan;29(1):212-219 [FREE Full text] [doi: 10.1002/pon.5256] [Medline: 31654533]

131. Compen FR, Bisseling EM, Schellekens MP, Jansen ET, van der Lee ML, Speckens AE. Mindfulness-based cognitive therapy for cancer patients delivered via internet: qualitative study of patient and therapist barriers and facilitators. J Med Internet Res 2017 Dec 18;19(12):e407 [FREE Full text] [doi: 10.2196/jmir.7783] [Medline: 29254912]

132. Boggs JM, Beck A, Felder JN, Dimidjian S, Metcalf CA, Segal ZV. Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: a qualitative study. J Med Internet Res 2014 Mar 24;16(3):e87 [FREE Full text] [doi: 10.2196/jmir.3129] [Medline: 24662625]

133. Mulder R, Murray G, Rucklidge J. Common versus specific factors in psychotherapy: opening the black box. Lancet Psychiatry 2017 Dec;4(12):953-962. [doi: 10.1016/S2215-0366(17)30100-1] [Medline: 28689019]

134. Brito G. Rethinking mindfulness in the therapeutic relationship. Mindfulness 2013 Jan 11;5(4):351-359. [doi: 10.1007/s12671-012-0186-2]

135. Cavanagh K, Millings A. (Inter)personal computing: the role of the therapeutic relationship in e-mental health. J Contemp Psychother 2013 Jul 17;43(4):197-206. [doi: 10.1007/s10879-013-9242-z]

136. Ormrod JA, Kennedy L, Scott J, Cavanagh K. Computerised cognitive behavioural therapy in an adult mental health service: a pilot study of outcomes and alliance. Cogn Behav Ther 2010;39(3):188-192. [doi: 10.1080/16506071003675614] [Medline: 20485996]

137. Clough BA, Casey LM. Technological adjuncts to increase adherence to therapy: a review. Clin Psychol Rev 2011 Jul;31(5):697-710. [doi: 10.1016/j.cpr.2011.03.006] [Medline: 21497153]

Abbreviations

AAF: ambulant activity feedback
ACT: acceptance and commitment therapy
eMBP: eHealth mindfulness-based program
ITT: intention-to-treat

MBCR: mindfulness-based cancer recovery
MBCT: mindfulness-based cognitive therapy
MBCT-Ca: mindfulness-based cognitive therapy for cancer
MBP: mindfulness-based program
MBSR: mindfulness-based stress reduction
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial

XSL•FO
RenderX

<u>Review</u>

Economic Value of Data and Analytics for Health Care Providers: Hermeneutic Systematic Literature Review

Philip von Wedel[1], MSc; Christian Hagist[1], Prof Dr

Chair of Economic and Social Policy, WHU - Otto Beisheim School of Management, Vallendar, Germany

Corresponding Author:
Philip von Wedel, MSc
Chair of Economic and Social Policy
WHU - Otto Beisheim School of Management
Burgplatz 2
Vallendar, 56179
Germany
Phone: 49 02616509 ext 255
Email: philip.wedel@whu.edu

Abstract

Background: The benefits of data and analytics for health care systems and single providers is an increasingly investigated field in digital health literature. Electronic health records (EHR), for example, can improve quality of care. Emerging analytics tools based on artificial intelligence show the potential to assist physicians in day-to-day workflows. Yet, single health care providers also need information regarding the economic impact when deciding on potential adoption of these tools.

Objective: This paper examines the question of whether data and analytics provide economic advantages or disadvantages for health care providers. The goal is to provide a comprehensive overview including a variety of technologies beyond computer-based patient records. Ultimately, findings are also intended to determine whether economic barriers for adoption by providers could exist.

Methods: A systematic literature search of the PubMed and Google Scholar online databases was conducted, following the hermeneutic methodology that encourages iterative search and interpretation cycles. After applying inclusion and exclusion criteria to 165 initially identified studies, 50 were included for qualitative synthesis and topic-based clustering.

Results: The review identified 5 major technology categories, namely EHRs (n=30), computerized clinical decision support (n=8), advanced analytics (n=5), business analytics (n=5), and telemedicine (n=2). Overall, 62% (31/50) of the reviewed studies indicated a positive economic impact for providers either via direct cost or revenue effects or via indirect efficiency or productivity improvements. When differentiating between categories, however, an ambiguous picture emerged for EHR, whereas analytics technologies like computerized clinical decision support and advanced analytics predominantly showed economic benefits.

Conclusions: The research question of whether data and analytics create economic benefits for health care providers cannot be answered uniformly. The results indicate ambiguous effects for EHRs, here representing data, and mainly positive effects for the significantly less studied analytics field. The mixed results regarding EHRs can create an economic barrier for adoption by providers. This barrier can translate into a bottleneck to positive economic effects of analytics technologies relying on EHR data. Ultimately, more research on economic effects of technologies other than EHRs is needed to generate a more reliable evidence base.

(J Med Internet Res 2020;22(11):e23315)   doi:10.2196/23315

Introduction

Data and analytics applications increasingly find their way into our health care systems. Some manifestations of data, like the electronic health record (EHR), have already been more established in many member countries of the Organisation of Economic Co-operation and Development. Analytics technologies such as computerized clinical decision support (CCDS) or advanced analytics (AA) based on big data and

XSL•FO
RenderX

artificial intelligence (AI) still seem to be newcomers in this field. Hopes are high that data and analytics significantly improve quality, efficiency, and patient experience of health care delivery [1]. Taking the perspective of health care systems, latest research, indeed, shows that adoption of EHRs leads to fewer medication errors, less adverse drug effects, and higher guideline adherence [2,3]. The use of clinical decision support (CDS) tools is associated with lower morbidity, potentially improving mortality [4,5]. Based on EHR data, AA is already able to predict the onset of several diseases like diabetes, schizophrenia, or cancer as well as provide care-related forecasts of in-hospital mortality, unplanned readmissions, length of stay, or infection risks [6-9]. One of the more recent topics is the possibility to diagnose the novel coronavirus disease (COVID-19) by applying AI to chest computed tomography scans [10,11]. It becomes clear how the introduction of these technologies can clearly create positive spillover effects for the entire health care system. When taking the microperspective of single providers, however, current adoption of data and analytics seems to paint a different picture in many countries. In the United States or Denmark, almost all hospitals work with a sophisticated EHR, while many European countries show much lower adoption rates. For example, reports indicate that 38.3% to 47.4% of German or 27.8% to 46.4% of Austrian hospitals lacked a system entirely in 2017 [12,13]. Even though analytics applications relying on AI or big data show strong potential, adoption in everyday provider operations is still comparatively low [14]. The reasons for this are manifold and include social, ethical, legal, or technological barriers [15]. The most powerful barrier, however, is still of an economic nature. Health care providers see the initial and ongoing maintenance costs as key barriers for adoption and oftentimes question overall cost-effectiveness of these solutions [15-17]. In a world that has not yet significantly pivoted towards quality-based reimbursement, quality improvements via data and analytics are, ironically, not necessarily directly linked to economic benefits for single health care providers. The much higher adoption of EHRs in the United States can, to a large part, be explained by strong financial subsidies by policy makers [18,19]. Taking the single provider's perspective, the question pertains whether hospitals, clinics, and practices can gain economic benefits from the usage of data and analytics. Most existing reviews in this field heavily focus on EHRs, but do not take into consideration other analytics tools [20-23]. Other more recent reviews focus on the economic impact of single areas of data and analytics like AI, but not specifically on providers [24]. Our work attempts to fill this gap by providing a comprehensive review of the literature covering the economic impact of several applications of data and analytics exclusively on providers. In the end, the promising potential of a number of established and rapidly evolving technologies to improve quality of care in our health care systems can only be optimally leveraged via widespread adoption by single providers.

Methods

Hermeneutic Systematic Review

The common systematic review ideally represents a highly structured approach for searching, screening, including, and summarizing studies to answer a rather narrowly defined question [25,26]. It might not, however, show perfect fit with all research questions. As Greenhalgh et al [27] summarized, it often "can be viewed as a set of methodologies characterized by tight focus, exhaustive search, high rejection-to-inclusion ratio and an emphasis on technical rather than interpretive synthesis methods." The hermeneutic review methodology introduced by Boell and Cecez-Kecmanovic [28] showed a particularly good fit with the broader nature of this study's research question. This process of a literature review follows 2 interlinked cycles: (1) search and acquisition and (2) analysis and interpretation (see Multimedia Appendix 1). The hermeneutic process allows and encourages a constant process of refining and extending the search realm of cycle (1) by deeply engaging with the content of the identified literature via cycle (2). This enables the researcher to leverage "the importance of reading and dialogical interaction between the literature and the researcher, […] seeking originality rather than replicability" [28]. Nevertheless, to assure the systematic execution of this review, guidelines for scoping studies including a 6-step process by Arksey and O'Malley [29] were followed. The hermeneutic approach was hereby complemented by the established tools for study identification and charting, assuring a systematic execution of the review. These tools resulted in a clear, reproducible, and structured overview of how studies were identified and for which reasons studies were excluded. In the end, by combining the traits of systematic and hermeneutic reviews, this study attempted to generate a structured, reproducible, comprehensive, and content-focused review of the literature.

Search Strategy

Literature included in this review was identified via iterative structured keyword searches in the online databases PubMed and Google Scholar, as well as a complementary backward and manual search. The following keyword search on article titles was applied to both databases: (x) AND (cost(s) OR revenue OR benefit OR return OR ROI OR value OR efficiency OR productivity) AND (hospital(s) OR practice(s) OR provider(s)). In this search, x represented a placeholder for terms that were iteratively added following the hermeneutic approach, and the following segments assured inclusion of studies only covering economic effects for only health care providers and which remained unchanged for all searches. In an initial search, x was comprised of "Electronic Health Record," "Electronic Medical Record," "Electronic Patient Record," "Analytics," and "Clinical Decision Support" (including all alternative and plural types of wording and abbreviations). Following the hermeneutic approach, both authors independently screened the resulting studies and jointly decided on additional search terms, expanding x to also include the terms "Algorithm," "Artificial Intelligence," "Big Data," "Machine Learning," "Deep Learning," "Natural Language Processing," and "Telemedicine." Interestingly, searches in the field of mobile health (mHealth) including health applications did not generate any suitable results. The search was limited to journal articles published in English between January 2009 and December 2019. The exact search queries for both databases can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Results

Study Selection

The PubMed and Google Scholar searches generated 79 and 165 results, respectively (see Figure 1). Following deduplication, a total of 165 studies remained for more detailed review. Only published journal articles were considered for the review. Hence, the deduplicated search results were again cleaned, resulting in 113 articles. Titles, abstracts, and, if needed, content of these articles were analyzed by both authors independently in order to determine fit to the research question, narrowing results to 43 results. Frequent reasons for exclusion were articles dealing with effects of data and analytics on stakeholders other than providers or on overarching national health care spending. Other examples were articles covering analogue tools and processes like paper-based decision support or diagnostic testing decision algorithms. An overview of all 113 screened studies and respective reasons for exclusion can be found in Multimedia Appendix 3. A complementary backward and manual search by both authors independently resulted in an additional 7 articles for inclusion. The final 50 articles were thoroughly reviewed, and key properties were summarized by the first author in a structured manner to facilitate pattern identification and final synthesis generation (see Multimedia Appendix 4).

Figure 1. Literature search and study selection process.
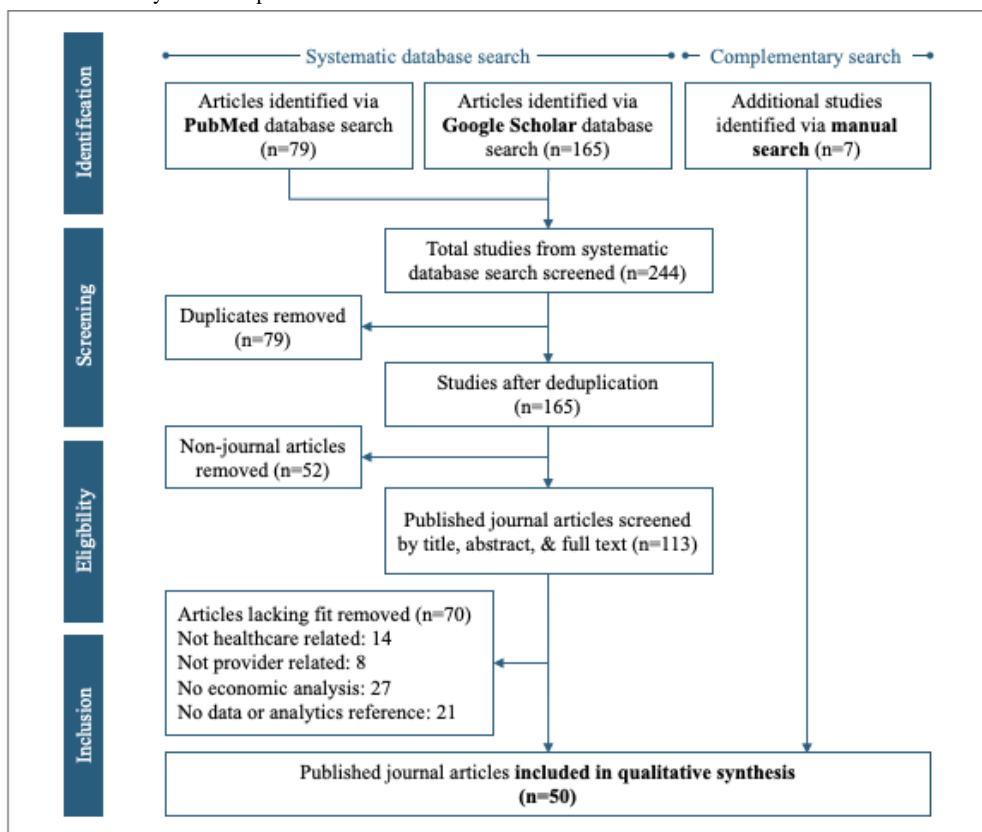
Study Categorization

Following in-depth review of the 50 final studies, 2 angles for categorization emerged. First, the studies were sorted according to the technology under research, resulting in 5 key categories, namely EHRs, CCDS, AA, business analytics (BA), and telemedicine. Second, studies were categorized based on the type of identified economic impact. This impact categorization consists of 2 combined components, namely mode (direct vs indirect) and direction (positive vs negative vs neutral vs mixed). Considering the impact mode, studies were categorized to have an *indirect* impact when no *direct* impact on costs or revenue but on efficiency or productivity was shown (see Figure 2 for a summary and details).

Figure 2. Overview of reviewed studies categorized by technology and economic impact.

Technology categorization	Description	Economic impact categorization (# studies)	+	−	=	≈
Electronic Health Records (EHR)	Studies examine the economic impact of EHRs in general without necessarily differentiating between single components (eg, computerized physician order entry, clinical decision support).	Total: 30				6
		Direct	8	2	-	
		Indirect	4	6	4	
Computerized Clinical Decision Support (CCDS)	Studies examine the economic impact of a CCDS tool such as laboratory order alerts that do not rely on advanced analytics.	Total: 8				-
		Direct	7	-	-	
		Indirect	1	-	-	
Advanced Analytics (AA)	Studies examine the economic impact of AA applications (eg, artificial intelligence, machine learning, deep learning) utilizing mainly health data from EHRs to support diagnosis, decision making, or workflows.	Total: 5				-
		Direct	-	-	-	
		Indirect	5	-	-	
Business Analytics (BA)	Studies examine the economic impact of BA tools mainly analyzing operational workflow or billing or claims data.	Total: 5				-
		Direct	5	-	-	
		Indirect	-	-	-	
Telemedicine	Studies examine the economic impact of telemedicine applications like remote monitoring.	Total: 2				-
		Direct	-	-	1	
		Indirect	1	-	-	

Economic impact: ⊕ Positive ⊖ Negative ⊜ Neutral ≈ Mixed   Direct: Cost and/or revenue impact; Indirect: Productivity and/or efficiency impact

Electronic Health Records

At 60% (30/50) of identified articles, EHRs represented the most comprehensive body of literature by far. In terms of economic impact, overall, a rather ambiguous pattern could be observed, with 12 studies revealing a positive, 8 studies a negative, 4 studies a neutral, and 6 studies a mixed economic effect on providers. The majority of studies was US-based (20/50, 40%), with Asia (3/60, 5%), the rest of the world (2/60, 3%), and Europe (0/60) showing less research activity. The remaining 5 articles represented international literature reviews.

The 5 literature reviews included in the sample predominantly indicated mixed economic impacts of EHRs. All reviews included studies proving positive effects mostly via increased efficiency; however, for almost every review, another identified article indicated the opposite [21-23,30]. Only Highfill [20] revealed overall positive economic effects, determining a 1.1%-13.8% cost decrease (95% CI) after EHR introduction in their meta-analysis.

Cost-benefit analyses (CBA) were presented in 5 articles and also painted a slightly ambiguous picture, with significantly varying timelines for EHR installations to break even. The majority of studies indicated an average breakeven timeline between 3 and 8 years for EHR implementations in hospitals [31-33]. Jang et al [34] indicated a much shorter 6.2-17.4–month (95% CI) breakeven timeframe for primary care clinics. Only 1 study revealed a clear negative impact, showing a negative 5-year return on investment [35]. In general, the CBAs provided some interesting practice-oriented insights for EHR implementations. Results from Choi et al [31] and

Adler-Milstein et al [35] both emphasized the importance of fully eliminating legacy costs like paper-based records and related dictation services. Parallel digital and analog structures resulted in fewer efficiency gains and, hence, longer breakeven timelines. Besides decreasing costs, a successful EHR introduction also focused on additional revenue generation via improved charge capture and reduction of billing errors [31,32,35]. Lastly, Jang et al [34] showed that more recent EHR systems and those using flow diagrams also came with shorter breakeven timelines, implying potential important technological advances by vendors over the years.

Besides the full CBAs, 6 studies examined the effects of EHR introductions on a variety of single-cost or revenue items. Encinosa and Bae [36] showed how the introduction of advanced EHRs reduced adverse drug effects from 3.6% to 1.4% of all cases, saving an average of US $4790 per avoided case. Joseph [37] revealed how personnel formerly needed for paper-based record keeping could be reduced, thereby saving more than US $6 million over 5 years. Zlabek et al [38] showed how transcription costs were significantly reduced, resulting in US $667,896 in costs saved 1 year after EHR introduction. A different source of cost savings was the avoidance of redundant laboratory tests and imaging exams. A computerized physician order entry (CPOE) system within an EHR resulted in an 18% decrease in laboratory test orders, as well as 6.3% fewer radiology exams [38,39]. However, Schnaus et al [40] revealed the importance of appropriate execution of a CPOE implementation. The authors examined a temporary change regarding the preselected laboratory test type when physicians searched for a complete blood count (CBC) within the CPOE

XSL•FO
RenderX

tool. For 23 days, the system preselected a slightly more costly version of a CBC. Presumably due to time constraints, a number of physicians did not double-check this preselected test type, which resulted in an average daily cost increase for CBC testing of US $293.10. Besides the direct economic impact via costs, some studies demonstrated positive effects via revenue. Terry [41] highlighted the potential from value-based reimbursement based on EHR data. The author saw the lack of an EHR system as "an 'opportunity cost' that can be quantified and weighed against the cost of installing a system" [41].

Finally, a significant share of papers (14/30) examined indirect economic impacts of EHRs via changes in efficiency or productivity. Here, a rather negative image emerged, with only 4 studies revealing positive effects, and the remaining showing either negative (6/30) or neutral (4/30) effects. Due to significant heterogeneity, it was difficult to draw generalizable insights from this sample of literature. This was a takeaway the identified systematic reviews also revealed. 5 studies examined productivity changes after EHR introduction, where productivity was mostly defined as average patient volumes. Of these studies, 3 revealed no statistically significant changes, hence neutral economic impacts [42-44]. Kaneko et al [45] showed a negative impact on multifactor productivity following EHR introduction in Japanese municipal hospitals. Only 1 study revealed positive long-term effects on productivity [46]. In 9 studies, efficiency implications were examined, where efficiency was defined rather heterogeneously as treatment times, waiting times, length of stay, or personnel volumes. While 1 study showed no effects [47], 5 studies revealed a negative economic impact [48-52]. Especially, the implementation of a fully-fledged EHR in a relatively short period of time, a so-called "big bang" introduction, seemed to be detrimental to hospital efficiency [51]. Only 3 studies showed somewhat limited positive effects on provider efficiency following EHR implementation [53-55].

Computerized Clinical Decision Support

Studies examining the economic impact of CCDS on providers represented the second-largest share of identified articles, at 16% (8/50). A strong picture regarding the impact emerged, with all 8 studies revealing a positive economic impact on providers, predominantly of a direct nature. Again, the majority of articles was US-based (5/8, 63%), and others were located in the rest of the world (2/8, 30%). The remaining article represented an international literature review.

Bright et al [5] presented the only included systematic review of CDS tools also assessing their impact on costs. Of a total 148 identified papers, 22 studies analyzed costs, of which 13 implied cost reductions. The authors saw this as "modest evidence from academic and community inpatient and ambulatory settings" [5]. Not all included studies, however, examined fully computerized CDS tools.

In 3 articles, it was shown how CCDS systems could reduce the number of imaging studies, laboratory tests, or the amount of medicine utilized. Fleddermann et al [56] assessed the introduction of an automated alert to avoid unnecessary ordering of echocardiography studies. Over the study period, 20% of the respective studies were cancelled, thereby saving the associated costs. Okumura et al [57] examined the cost savings associated

with implementing a tool to optimize antibiotic use in surgical prophylaxis. By reminding physicians of common standards of care, the system decreased the usage significantly by 1.26 defined daily doses per 100 bed days to –0.2 defined daily doses per 100 bed days (95% CI), thereby saving an estimated US $50,000 per 100 bed days. Lastly, Levick et al [58] assessed an alert for B-type natriuretic peptide testing. Again, the alert resulted in a test reduction of 21%, saving an estimated US $92,000 per year.

Besides effects via reduced volumes in tests or studies, 3 other articles revealed cost savings via supporting decisions regarding care processes and workflows. Quadros et al [59] examined CDS that supported fast tracking the discharge of certain patients after brain tumor surgery. The tool resulted in a significant length of stay reduction of 2 days on average, saving US $630 per hospitalization. Collins et al [60] showed how decisions on the timing of nasal feeding tube insertions for poststroke patients with dysphagia supported by CDS reduced the number of nasal tube replacements and repeat x-rays and the associated costs. It is important to mention here, however, that these 2 papers did not reveal whether the CDS tools were fully computerized. Lastly, Wagholikar et al [61] presented the impact of a CCDS tool in an outpatient setting. Here, the tool supported physicians with chart review via a computerized checklist to decide on preventive services and management of chronic diseases. The tool showed an indirect positive economic impact by reducing review times by 65% per patient.

The eighth article in the CCDS category by Elkin et al [62] is the only one examining direct cost savings based on supporting diagnosis. The authors applied a differential diagnosis support tool to cases in diagnostically challenging Diagnostic Related Groups and found that, for these patients, the provider costs per case were reduced by 3.7%, to 19.5% (95% CI).

Advanced Analytics

The recently increasingly prominent field of AA including AI, machine learning, and deep learning represented only 10% of the identified literature (5/50). A very homogenous picture was painted, with all 5 studies indicating indirect positive economic effects on providers; 3 articles were US-based, and others originated from Europe (1/5) and Asia (1/5).

The identified articles showed 2 main use cases of AA. First, 2 articles showed how AA could support decision making in the field of imaging. Lee [63] applied a convolutional neural network to determine musculoskeletal magnetic resonance imaging scanning protocols. The authors hypothesized that this assistance in protocol generation could potentially save personnel time and hence improve provider efficiency. The second article presented the deployment of the IBM Watson natural language processing model to automatically decide on the usage of intravenous contrast for magnetic resonance imaging protocols [64]. Again, the authors hypothesized that this support in decision making has the potential to drive provider efficiency.

The second use case represented the prediction of patients' disease progression and the associated care processes. Wang et al [65] showed how a convolutional neural network–based tool

XSL•FO

RenderX

using hospital EHR data could predict readmissions. Readmission predictions can be valuable information since a majority of readmissions is associated with penalties for providers. Nevertheless, the authors only hypothesized this potential benefit. Almeida [66] presented a case study of a hospital center in Portugal that applied a big data analytics platform. Based on EHR and vital sign data, the system was able to correctly predict 30% of intensive care unit admissions and 50% of non-intensive care unit inpatient deaths. Again, the author hypothesized potential efficiency improvements. Lastly, Peck et al [67] represented the only article in the AA category, which proved actual efficiency improvements instead of only hypothesizing them. The authors presented the impact of a tool predicting the patient flow from the emergency department to the inpatient units via discrete event simulation. By sharing information on crowding levels and total expected beds needed with physicians and nurses, the boarding time from the emergency department to inpatient units was reduced by between 11.69% and 18.38%, depending on hospital type.

Business Analytics

Studies examining the economic impact of BA on providers represented 10% of identified articles (5/50). All 5 studies revealed a positive economic effect on providers, mostly of a direct nature. The majority of articles represented US-based (4/5) or Europe-based (1/5) case studies.

BA tools analyzing equipment utilization were examined in 2 articles. Stekel et al [68] examined the example of an ultrasound practice that used probe utilization data to support purchasing decisions. The analysis of procedure data resulted in the decision to not replace a broken probe, thereby saving US $10,000. Swedberg [69] showed how attaching radiofrequency identification tags to all equipment in a 1100-bed hospital increased equipment utilization rates from 5% to 40%. The system was able to reduce the need to rent or purchase additional equipment, saving an estimated US $200,000 per year.

Examples of how BA applications improved billing by reducing revenue leakage or avoiding penalties were presented in 3 articles [70-72]. For example, Dulac et al [70] showed in a case study how a US-based hospital used data analytics to uncover root causes for an increase in preventable complications and readmission rates that implied payment reductions totaling 3.5% of total revenue. The tool helped the hospital to ultimately reduce penalties to 0%.

Telemedicine

Studies examining the economic impact of telemedicine on providers represented the smallest share of identified articles (2/50). Both articles originated from Europe, with 1 study revealing an indirect positive effect and 1 a neutral effect.

Stoves et al [73] examined the advantages of an electronic medical round connecting general practitioners with nephrologists in the field of chronic kidney care. The program was perceived to improve efficiency for both physicians and nephrologists, as indicated in interviews and questionnaires; however, no quantitative efficiency improvements were reported. Heidbuchel et al [74], on the other hand, specifically analyzed differences in costs and financial impact between

remote and in-office follow-ups for implantable cardiac defibrillators. On average, the total cost and net financial impact for providers were neutral, not showing differences between remote and in-office follow-ups. Importantly, however, regional heterogeneity could be observed where providers in countries with remote follow-up reimbursement in places like Germany maintained or improved economics.

Discussion

Principal Findings

At first indication, the presented results appear to generate an overall positive answer to the overarching research question of the economic impact data and analytics have on health care providers. Of the 50 reviewed articles, 31 indicated a positive impact either via direct cost or revenue effects or via efficiency or productivity improvements. Studies showed how EHRs can, for example, directly save storage and personnel costs associated with paper records or increase physician productivity by making information available when and where it is needed. Other studies proved that CCDS can save material and labor costs by avoiding redundant laboratory tests and imaging studies. A more nuanced look at the results, however, shows that it is very important to differentiate between the 5 identified technology categories. In line with other literature reviews, a mixed overall picture, at best, was revealed for the economic impact of EHRs, or "data," on providers. From a provider's perspective, 12 studies revealing a positive result and 18 revealing negative, neutral, or mixed results do not necessarily promote a quick decision on EHR investments, at least from an economic point of view. On the other hand, "analytics" applications like CCDS, AA, and BA seem to predominantly generate positive economic effects. Nevertheless, the small number of identified papers covering these technologies, yet again, points at the risk of discouraging rapid adoption by providers from an economic point of view. Ultimately, this review reveals a rather uncomfortable decision-making situation for providers with the economic impact of "data," represented by EHRs, being exhaustively researched but revealing ambiguous results and "analytics" indicating positive results but being only sparsely investigated.

Considering the positive effects of EHRs on health care outcomes, the identified ambiguous results regarding the economic impact for providers also implies potentially missing out on the associated welfare gains across populations. Some nations' policy makers already acknowledged this and incentivize EHR adoption. The United States Health Information Technology for Economic and Clinical Health Act from 2009 injected several billions of dollars into the system for subsidized EHR installations [19]. This approach seemed to have worked as intended by pushing EHR adoption closer to 100% for US hospitals [18]. Of course, national health systems strongly vary, but this outcome should at least foster a discussion of whether EHR subsidization might also be a solution in other countries with comparably low current EHR adoption rates. Germany, for example, announced a "hospital future law" (Krankenhauszukunftsgesetz) as part of a COVID-19 stimulus package in 2020, which envisions an up to €4.3 billion fund for investments in digital infrastructures and emergency capacities

[75]. Here, it is important to point out the positive spillover effects of data and analytics for the entire health care system again. Even though this review predominantly takes a microperspective of the single provider, in the end, adoption on the microlevel is a key prerequisite for changes or improvements at the system level. Data and analytics might provide proven positive effects on quality of care, but until the world does significantly pivot towards quality-based care and reimbursement, alternative ways to foster technology adoption should be considered.

Another effect involving EHRs is also important to consider. Several of the included studies showed how EHR installations can act as a door-opener to other technologies that actually seem to predominantly have positive economic effects for the provider. Especially, 2 other technology categories identified in this review, namely CCDS and AA, strongly rely on data contained in EHRs. More precisely, 75% (6/8) of the CCDS tools and 60% (3/5) of the AA tools in the identified studies needed some sort of EHR input. Even if somewhat limited in quantity, the current identified research in these 2 fields revealed only positive economic effects for providers. Hence, EHR adoption can become a bottleneck to the positive economic effects of technologies further down the line like CCDS and AA. Following EHR adoption, providers are likely able to derive economic benefits from adjoined technologies identified in this review. More research dedicated to these economic effects of supplementing EHRs with adjoined technologies like CCDS and AA is needed to derive a more targeted evidence base.

Leaving policy implications aside, our work generates insights for providers as well. For providers considering an EHR installation, this review showed important factors for an economically feasible introduction. Eliminating all legacy costs like paper-based records and related dictation services, repurposing paper record space into clinical space, or installing new technology in a stepwise fashion (avoiding a big bang) are all important takeaways from this review. For hospitals or practices already using an EHR, adjacent technologies, like CCDS or AA, can provide economic benefits, potentially even resulting in a shorter breakeven time for the EHR installation. Additionally, the emerging opportunities to participate in value-based care plans utilizing EHR data or the utilization of business intelligence should not be fully neglected. Nevertheless, a number of other potential sources of economic value from data seems not to be currently covered by research. For example, no study was identified that covered the potential for direct monetization of anonymized patient data or the ability to drive patient volumes by marketing the application of advanced digital tools.

Ultimately, it is important to note that research on the economic impact of data and analytics on providers remains rather limited in geographies other than the United States and in technologies other than EHR. In general, this review did not identify geography as a predictor for the type of economic impact. However, with almost 65% (32/50) of included articles being US-based, more research in other geographies is needed to draw a definite conclusion whether geographies and related health care systems are significant drivers. From a technology perspective, the few studies covering technologies other than EHR revealed proof for economic advantages; however, no comprehensive cost-benefit analyses and few systematic reviews were identified for these technologies. In the field of AA, 80% (4/5) of identified studies only hypothesized economic benefits. In the near future, however, vendors of AA tools need to also provide high-quality proof for the economic advantages of their solutions.

Limitations

This work is exposed to limitations that are mostly inherent to literature reviews in general. Only PubMed and Google Scholar online databases were searched; hence, relevant research captured exclusively by other databases could have been excluded. The sample of identified articles potentially lacks certain avenues of research not captured by the structured keyword search, thereby missing other technologies. The applied hermeneutic systematic search approach, however, worked against these limitations by explicitly allowing for iterative searches. Additionally, the systematic search was complemented by manual search techniques. On a different note, most identified studies are based in the United States; hence, conclusions might not be fully applicable to other geographies. Ultimately, it is important to note that the research subject "data" is almost exclusively represented by studies focusing on EHRs, thereby not touching on other potentially relevant sources and applications of data. Nevertheless, EHRs can be considered as a key data container in the context of health care. The research subject "analytics," on the other hand, faces a very limited body of evidence, which strongly impacts the generalizability of this study's findings. More research covering these other technologies is needed to generate a more holistic and reliable evidence base. Lastly, the intended broad spectrum of reviewed studies prevents a clear and uniform definition and quantification of "economic value." Studies and respective results can, hence, not be compared on the same scale, also since the methodological quality of the original studies was not analyzed.

Conclusion

This review synthesized literature examining the economic value of data and analytics for health care providers. Five key technologies were identified, namely EHRs, CCDS, AA, BA, and telemedicine. Overall, 31 of the 50 reviewed articles indicated a positive economic impact, either via direct cost or revenue effects or via efficiency or productivity improvements. A more nuanced view showed that this is especially the case for less studied technologies like CCDS, AA (including AI and big data analysis), and BA. For the most extensively studied technology of EHRs, a more ambiguous view with varying economic impacts emerged. Since technologies like CCDS and AA strongly rely on EHR data, these ambiguous research findings have the potential to turn EHR adoption into a bottleneck for the adjoined technologies with mostly positive economic effects. This review also encourages discussions around how subsidization of EHRs, like that implemented in the United States and planned for Germany, could potentially unlock the proven economic potential of second-order adjoined technologies. It can be concluded that more research covering the economic effects of technologies other than EHRs would

significantly improve the current evidence base and potentially drive adoption by health care providers.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The interlinked cycles of the hermeneutic process.
[PNG File , 284 KB - jmir_v22i11e23315_app1.png]

Multimedia Appendix 2
Detailed search queries used for the PubMed and Google Scholar database searches.
[PDF File (Adobe PDF File), 113 KB - jmir_v22i11e23315_app2.pdf]

Multimedia Appendix 3
Overview of all published articles screened by title, abstract, and text (including reasons for exclusion).
[PDF File (Adobe PDF File), 286 KB - jmir_v22i11e23315_app3.pdf]

Multimedia Appendix 4
Detailed syntheses of the final 50 studies included in this review.
[PDF File (Adobe PDF File), 332 KB - jmir_v22i11e23315_app4.pdf]

References
1. Blumenthal D, Glaser JP. Information Technology Comes to Medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [doi: 10.1056/nejmhpr066212]
2. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb 30;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
3. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
4. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of Computerized Decision Support Systems Linked to Electronic Health Records: A Systematic Review and Meta-Analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/ajph.2014.302164]
5. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
6. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Med Inform 2016 Sep 30;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
7. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
8. Oh J, Makar M, Fusco C, McCaffrey R, Rao K, Ryan EE, et al. A Generalizable, Data-Driven Approach to Predict Daily Risk of Clostridium difficile Infection at Two Large Academic Health Centers. Infect Control Hosp Epidemiol 2018 Apr 26;39(4):425-433 [FREE Full text] [doi: 10.1017/ice.2018.16] [Medline: 29576042]
9. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
10. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020 Aug 19;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
11. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. The Lancet Digital Health 2020 Apr;2(4):e166-e167. [doi: 10.1016/s2589-7500(20)30054-6]
12. Stephani V, Busse R, Geissler A. Benchmarking der Krankenhaus-IT: Deutschland im internationalen Vergleich. In: Krankenhaus-Report 2019. Berlin: Springer Publishing Company; 2019:17-32.

13. Hübner U, Esdar M, Hüsers J, Liebe JD, Naumann L, Thye J, et al. IT-Report Gesundheitswesen. Osnabrück University of Applied Sciences. 2019 Dec 31. URL: https://www.hs-osnabrueck.de/fileadmin/HSOS/Homepages/IT-Report_Gesundheitswesen/IT-Report_Gesundheitswesen_2020_final.pdf [accessed 2020-11-12]

14. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

15. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]

16. Jamoom EW, Patel V, Furukawa MF, King J. EHR adopters vs. non-adopters: Impacts of, barriers to, and federal initiatives for EHR adoption. Healthc (Amst) 2014 Mar;2(1):33-39 [FREE Full text] [doi: 10.1016/j.hjdsi.2013.12.004] [Medline: 26250087]

17. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to Electronic Health Record Adoption: a Systematic Literature Review. J Med Syst 2016 Dec;40(12):252 [FREE Full text] [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]

18. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422 [FREE Full text] [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]

19. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]

20. Highfill T. Do hospitals with electronic health records have lower costs? A systematic review and meta-analysis. International Journal of Healthcare Management 2019 May 21;13(1):65-71. [doi: 10.1080/20479700.2019.1616895]

21. Rahmawati SR. Could We Derive Benefit From Implementing Electronic Medical Records In Hospital?: A Structured Evidence and Narrative Review. arkesmas 2019 Mar 16;3(2):56-71. [doi: 10.22236/arkesmas.v3i2.3008]

22. Reis ZSN, Maia TA, Marcolino MS, Becerra-Posada F, Novillo-Ortiz D, Ribeiro ALP. Is There Evidence of Cost Benefits of Electronic Medical Records, Standards, or Interoperability in Hospital Information Systems? Overview of Systematic Reviews. JMIR Med Inform 2017 Aug 29;5(3):e26 [FREE Full text] [doi: 10.2196/medinform.7400] [Medline: 28851681]

23. Uslu A, Stausberg J. Value of the Electronic Medical Record for Hospital Care: A Review of the Literature. Journal of Healthcare Engineering 2011 Sep;2(3):271-284. [doi: 10.1260/2040-2295.2.3.271]

24. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. J Med Internet Res 2020 Feb 20;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]

25. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: John Wiley & Sons; 2008.

26. Institute of Medicine. In: Eden J, Levit L, Berg A, Morton S, editors. Finding what works in health care: standards for systematic reviews. Washington, D.C: National Academic Press; 2011.

27. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest 2018 Jun 16;48(6):e12931 [FREE Full text] [doi: 10.1111/eci.12931] [Medline: 29578574]

28. Boell SK, Cecez-Kecmanovic D. A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. CAIS 2014;34:12. [doi: 10.17705/1cais.03412]

29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

30. Thompson G, O'Horo JC, Pickering BW, Herasevich V. Impact of the Electronic Medical Record on Mortality, Length of Stay, and Cost in the Hospital and ICU: A Systematic Review and Metaanalysis. Crit Care Med 2015 Jun;43(6):1276-1282. [doi: 10.1097/CCM.0000000000000948] [Medline: 25756413]

31. Choi JS, Lee WB, Rhee P. Cost-benefit analysis of electronic medical record system at a tertiary care hospital. Healthc Inform Res 2013 Sep;19(3):205-214 [FREE Full text] [doi: 10.4258/hir.2013.19.3.205] [Medline: 24175119]

32. Li K, Naganawa S, Wang K, Li P, Kato K, Li X, et al. Study of the cost-benefit analysis of electronic medical record systems in general hospital in China. J Med Syst 2012 Oct 3;36(5):3283-3291. [doi: 10.1007/s10916-011-9819-6] [Medline: 22212632]

33. Zimlichman E, Keohane C, Franz C, Everett WL, Seger DL, Yoon C, et al. Return on Investment for Vendor Computerized Physician Order Entry in Four Community Hospitals: The Importance of Decision Support. The Joint Commission Journal on Quality and Patient Safety 2013 Jul;39(7):312-318. [doi: 10.1016/s1553-7250(13)39044-8]

34. Jang Y, Lortie MA, Sanche S. Return on investment in electronic health records in primary care practices: a mixed-methods study. JMIR Med Inform 2014 Sep 29;2(2):e25 [FREE Full text] [doi: 10.2196/medinform.3631] [Medline: 25600508]

35. Adler-Milstein J, Green CE, Bates DW. A survey analysis suggests that electronic health records will yield revenue gains for some practices and losses for many. Health Aff (Millwood) 2013 Mar;32(3):562-570. [doi: 10.1377/hlthaff.2012.0306] [Medline: 23459736]

36. Encinosa WE, Bae J. Will meaningful use electronic medical records reduce hospital costs? Am J Manag Care 2013 Nov;19(10 Spec No):eSP19-eSP25 [FREE Full text] [Medline: 24511884]

37. Joseph M. Meaningful streamlining. Hybrid practice management/EMR system boosts quality, reduces costs. Mod Healthc 2010 Jul 05;40(27):C8. [Medline: 20666214]

38. Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. J Am Med Inform Assoc 2011 Mar 01;18(2):169-172 [FREE Full text] [doi: 10.1136/jamia.2010.007229] [Medline: 21292703]

39. Silow-Carroll S, Edwards JN, Rodin D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. Issue Brief (Commonw Fund) 2012 Jul;17:1-40. [Medline: 22826903]

40. Schnaus MJ, Michalik M, Skarda P. Effects of Electronic Medical Record Display on Provider Ordering Behavior: Leveraging the EMR to Improve Quality and Costs. Am J Med 2017 Dec;130(12):1366-1371. [doi: 10.1016/j.amjmed.2017.08.019] [Medline: 28888466]

41. Terry K. Value-based incentives can help practices offset EHR costs. Med Econ 2014 May 10;91(9):41-43. [Medline: 25233781]

42. Dandu N, Zmistowski B, Chen AF, Chapman T, Howley M. How are Electronic Health Records Associated with Provider Productivity and Billing in Orthopaedic Surgery? Clinical Orthopaedics and Related Research 2019;477(11):2443-2451. [doi: 10.1097/corr.0000000000000896]

43. Redd TK, Read-Brown S, Choi D, Yackel TR, Tu DC, Chiang MF. Electronic health record impact on productivity and efficiency in an academic pediatric ophthalmology practice. J AAPOS 2014 Dec;18(6):584-589 [FREE Full text] [doi: 10.1016/j.jaapos.2014.08.002] [Medline: 25456030]

44. Voleti VB, Braunstein AL, Mahabir C, Schrier A, Chiang MF. Evaluation Of An Academic Eye Clinic's Practice Efficiency During The Early Transition Period To An Electronic Health Record System. Investigative Ophthalmology & Visual Science 2011;52(14):5549-5549.

45. Kaneko K, Onozuka D, Shibuta H, Hagihara A. Impact of electronic medical records (EMRs) on hospital productivity in Japan. Int J Med Inform 2018 Oct;118:36-43. [doi: 10.1016/j.ijmedinf.2018.07.008] [Medline: 30153919]

46. De Leon S, Connelly-Flores A, Mostashari F, Shih SC. The business end of health information technology. Can a fully integrated electronic health record increase provider productivity in a large community practice? J Med Pract Manage 2010;25(6):342-349. [Medline: 20695245]

47. Kadish SS, Mayer EL, Jackman DM, Pomerantz M, Brady L, Dimitriadis A, et al. Implementation to Optimization: A Tailored, Data-Driven Approach to Improve Provider Efficiency and Confidence in Use of the Electronic Medical Record. JOP 2018 Jul;14(7):e421-e428. [doi: 10.1200/jop.18.00093]

48. Chuang T, Yii N, Nyandowe M, Iyer R. Examine the impact of the implementation of an electronic medical record system on operating theatre efficiency at a teaching hospital in Australia. Int Surg J 2019 Apr 29;6(5):1453. [doi: 10.18203/2349-2902.isj20191865]

49. Furukawa MF, Raghu TS, Shao BBM. Electronic medical records and cost efficiency in hospital medical-surgical units. Inquiry 2010 Jun;47(2):110-123 [FREE Full text] [doi: 10.5034/inquiryjrnl_47.02.110] [Medline: 20812460]

50. Hollenbeck SM, Bomar JD, Wenger DR, Yaszay B. Electronic Medical Record Adoption. Journal of Pediatric Orthopaedics 2017;37(6):424-428. [doi: 10.1097/bpo.0000000000000679]

51. Huerta TR, Thompson MA, Ford EW, Ford WF. Electronic health record implementation and hospitals' total factor productivity. Decision Support Systems 2013 May;55(2):450-458. [doi: 10.1016/j.dss.2012.10.004]

52. McDowell J, Wu A, Ehrenfeld JM, Urman RD. Effect of the Implementation of a New Electronic Health Record System on Surgical Case Turnover Time. J Med Syst 2017 Mar 27;41(3):42. [doi: 10.1007/s10916-017-0690-y] [Medline: 28130725]

53. Furukawa MF. Electronic medical records and the efficiency of hospital emergency departments. Med Care Res Rev 2011 Feb 16;68(1):75-95. [doi: 10.1177/1077558710372108] [Medline: 20555014]

54. Kazley AS, Ozcan YA. Electronic medical record use and efficiency: A DEA and windows analysis of hospitals. Socio-Economic Planning Sciences 2009 Sep;43(3):209-216. [doi: 10.1016/j.seps.2008.10.001]

55. Pyron L, Carter-Templeton H. Improved Patient Flow and Provider Efficiency After the Implementation of an Electronic Health Record. CIN: Computers, Informatics, Nursing 2019;37(10):513-521. [doi: 10.1097/cin.0000000000000553]

56. Fleddermann A, Jones S, James S, Kennedy KF, Main ML, Austin BA. Implementation of Best Practice Alert in an Electronic Medical Record to Limit Lower-Value Inpatient Echocardiograms. Am J Cardiol 2018 Nov 01;122(9):1574-1577. [doi: 10.1016/j.amjcard.2018.07.017] [Medline: 30172364]

57. Okumura LM, Veroneze I, Burgardt CI, Fragoso MF. Effects of a computerized provider order entry and a clinical decision support system to improve cefazolin use in surgical prophylaxis: a cost saving analysis. Pharm Pract (Granada) 2016 Sep 30;14(3):717. [doi: 10.18549/pharmpract.2016.03.717]

58. Levick DL, Stern G, Meyerhoefer CD, Levick A, Pucklavage D. "Reducing unnecessary testing in a CPOE system through implementation of a targeted CDS intervention". BMC Med Inform Decis Mak 2013 Apr 08;13(1):43 [FREE Full text] [doi: 10.1186/1472-6947-13-43] [Medline: 23566021]

59. Quadros DG, Neville IS, Urena FM, Solla DFJ, Paiva WS, Teixeira M. Safety and Costs Analysis of a Fast-track Algorithm for Early Hospital Discharge After Brain Tumor Surgery. Neurosurgery 2019;66(Supplement_1):nyz310_432. [doi: 10.1093/neuros/nyz310_432]

60. Collins LB, Ward CW, Boggs B. Abstract WP492: Dysphagia Severity and Decision Making Algorithm Impact on Length of Hospital Stay, Restraint Use and Cost in Stroke Patients. Stroke 2019 Feb;50(Suppl_1):AWP492. [doi: 10.1161/str.50.suppl_1.wp492]

61. Wagholikar KB, Hankey RA, Decker LK, Cha SS, Greenes RA, Liu H, et al. Evaluation of the effect of decision support on the efficiency of primary care providers in the outpatient practice. J Prim Care Community Health 2015 Jan;6(1):54-60 [FREE Full text] [doi: 10.1177/2150131914546325] [Medline: 25155103]

62. Elkin PL, Liebow M, Bauer BA, Chaliki S, Wahner-Roedler D, Bundrick J, et al. The introduction of a diagnostic decision support system (DXplain™) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). Int J Med Inform 2010 Nov;79(11):772-777 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.09.004] [Medline: 20951080]

63. Lee YH. Efficiency Improvement in a Busy Radiology Practice: Determination of Musculoskeletal Magnetic Resonance Imaging Protocol Using Deep-Learning Convolutional Neural Networks. J Digit Imaging 2018 Oct 4;31(5):604-610 [FREE Full text] [doi: 10.1007/s10278-018-0066-y] [Medline: 29619578]

64. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH. Automatic Determination of the Need for Intravenous Contrast in Musculoskeletal MRI Examinations Using IBM Watson's Natural Language Processing Algorithm. J Digit Imaging 2018 Apr;31(2):245-251 [FREE Full text] [doi: 10.1007/s10278-017-0021-3] [Medline: 28924815]

65. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting Hospital Readmission via Cost-Sensitive Deep Learning. IEEE/ACM Trans. Comput. Biol. and Bioinf 2018 Nov 1;15(6):1968-1978. [doi: 10.1109/tcbb.2018.2827029]

66. Almeida JP. A disruptive Big data approach to leverage the efficiency in management and clinical decision support in a Hospital. Porto Biomed J 2016;1(1):40-42 [FREE Full text] [doi: 10.1016/j.pbj.2015.12.001] [Medline: 32258546]

67. Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. Characterizing the value of predictive analytics in facilitating hospital patient flow. IIE Transactions on Healthcare Systems Engineering 2014 Aug 18;4(3):135-143. [doi: 10.1080/19488300.2014.930765]

68. Stekel SF, Long Z, Tradup DJ, Hangiandreou NJ. Use of Image-Based Analytics for Ultrasound Practice Management and Efficiency Improvement. J Digit Imaging 2019 Apr 8;32(2):251-259 [FREE Full text] [doi: 10.1007/s10278-018-0151-2] [Medline: 30623273]

69. Swedberg C. RFID boosts medical equipment usage at UK hospital. RFID Journal. 2013 Aug 15. URL: https://www.rfidjournal.com/rfid-boosts-medical-equipment-usage-at-u-k-hospital#sthash.j8t6MAio.dpuf [accessed 2020-11-12]

70. Dulac J, Pryor R, Morrissey W. A data-driven approach to improving clinical care and reducing costs the experiences of an academic health center and a hospital system in applying data analytics provide valuable lessons for other organizations that are just beginning such efforts, which are critically important for success under value-based payment. Healthcare Financial Management 2017;71(8):52-58.

71. Rivera J, Delaney S. Using business analytics to improve outcomes: Orlando Health is taking advantage of customized solutions to get its hospital and physician practice revenue cycle systems in synch and strengthen its organization operationally and financially. Healthcare Financial Management 2015;69(2):64-68.

72. Schouten P. Big data in health careolving provider revenue leakage with advanced analytics. Healthcare Financial Management 2013;67(2):40-43.

73. Stoves J, Connolly J, Cheung CK, Grange A, Rhodes P, O'Donoghue D, et al. Electronic consultation as an alternative to hospital referral for patients with chronic kidney disease: a novel application for networked electronic health records to improve the accessibility and efficiency of healthcare. Qual Saf Health Care 2010 Oct 16;19(5):e54-e54. [doi: 10.1136/qshc.2009.038984] [Medline: 20554576]

74. Heidbuchel H, Hindricks G, Broadhurst P, Van Erven L, Fernandez-Lozano I, Rivero-Ayerza M, et al. EuroEco (European Health Economic Trial on Home Monitoring in ICD Patients): a provider perspective in five European countries on costs and net financial impact of follow-up with or without remote monitoring. Eur Heart J 2015 Jan 14;36(3):158-169 [FREE Full text] [doi: 10.1093/eurheartj/ehu339] [Medline: 25179766]

75. Bundesministerium für Gesundheit. Krankenhauszukunftsgesetz für die Digitalisierung von Krankenhäusern. 2020. URL: https://www.bundesgesundheitsministerium.de/krankenhauszukunftsgesetz.html [accessed 2020-09-18]

Abbreviations

AA: advanced analytics
AI: artificial intelligence
BA: business analytics
CBA: cost-benefit analysis
CBC: complete blood count
CCDS: computerized clinical decision support
CDS: clinical decision support
CPOE: computerized physician/provider order entry
EHR: electronic health record

XSL•FO
RenderX

XSL•FO
RenderX

Review

Mobile Apps to Reduce Tobacco, Alcohol, and Illicit Drug Use: Systematic Review of the First Decade

Petra Karin Staiger[1,2], PhD; Renee O'Donnell[3], PhD; Paul Liknaitzky[1], PhD; Rachel Bush[1], PhD; Joanna Milward[4], PhD

[1]School of Psychology, Deakin University, Faculty of Health, Geelong, Australia

[2]Centre for Drug Use, Addictive and Anti-social behaviour Research (CEDAAR), Deakin University, Geelong, Australia

[3]Monash Centre of Health Research and Implementation, Monash University, Clayton, Australia

[4]Department of Addictions, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, United Kingdom

Corresponding Author:
Petra Karin Staiger, PhD
School of Psychology
Deakin University
Faculty of Health
Locked Bag 20000
Geelong, 3220
Australia
Phone: 61 3 9244 6876
Fax: 61 3 92446858
Email: petra.staiger@deakin.edu.au

Abstract

Background: Mobile apps for problematic substance use have the potential to bypass common barriers to treatment seeking. Ten years following the release of the first app targeting problematic tobacco, alcohol, and illicit drug use, their effectiveness, use, and acceptability remains unclear.

Objective: This study aims to conduct a systematic literature review of trials evaluating mobile app interventions for problematic tobacco, alcohol, and illicit drug use.

Methods: The review was conducted according to recommended guidelines. Relevant databases were searched, and articles were included if the mobile app study was a controlled intervention trial and reported alcohol, tobacco, or illicit drug consumption as outcomes.

Results: A total of 20 studies met eligibility criteria across a range of substances: alcohol (n=11), tobacco (n=6), alcohol and tobacco (n=1), illicit drugs (n=1), and illicit drugs and alcohol (n=1). Samples included the general community, university students, and clinical patients. The analyzed intervention sample sizes ranged from 22 to 14,228, and content was considerably diverse, from simple stand-alone apps delivering self-monitoring or psychoeducation to multicomponent apps with interactive features and audio content, or used as adjuncts alongside face-to-face treatment. Intervention duration ranged from 1 to 35 weeks, with notifications ranging from none to multiple times per day. A total of 6 of the 20 app interventions reported significant reductions in substance use at post or follow-up compared with a comparison condition, with small to moderate effect sizes. Furthermore, two other app interventions reported significant reductions during the intervention but not at post treatment, and a third reported a significant interaction of two app intervention components.

Conclusions: Although most app interventions were associated with reductions in problematic substance use, less than one-third were *significantly* better than the comparison conditions at post treatment. A total of 5 out of the 6 apps that reported intervention effects targeted alcohol (of those, one targeted alcohol and illicit drugs and another alcohol and tobacco) and 1 targeted tobacco. Moreover, 3 out of 6 apps included feedback (eg, personalized) and 2 had high risk of bias, 1 some risk, and 3 low risk. All 6 apps included interventions of 6 weeks or longer. Common study limitations were small sample sizes; risk of bias; lack of relevant details; and, in some cases, poorly balanced comparison conditions. Appropriately powered trials are required to understand which app interventions are most effective, length of engagement required, and subgroups most likely to benefit. In sum, evidence to date for the effectiveness of apps targeting problematic substance use is not compelling, although the heterogeneous comparison conditions and trial designs across studies limit the ability to compare efficacy between apps. We discuss potential approaches that can help ascertain whether the promise of mobile app interventions for problematic substance use can be fulfilled.

Introduction

The problematic use of substances such as alcohol, tobacco, and illicit drugs is one of the leading causes of morbidity and mortality worldwide [1]. Despite the devastating health and social consequences, a large proportion of individuals who engage in problematic substance use do not seek formal treatment [2-4]. Help-seeking barriers include concern about anonymity, not knowing about or being able to access treatment, and the financial or time burdens of treatment [5-8]. Hence, interventions that can address some of these help-seeking barriers warrant attention to reduce the substantial negative impact of substances at a population level.

Mobile health (mHealth) interventions purport to overcome many of these help-seeking barriers by offering a population-based approach, improving access and affordability [9,10]. mHealth refers to health support delivered on mobile devices, such as cell phones, smartphones, and tablets [11], typically using dedicated apps, but also includes systems such as interactive voice response (IVR) and text messaging (SMS) [12,13]. Apps have rapidly become the most popular software method for delivering health support (ie, mHealth) on mobile devices. As of August 2019, a search of iTunes and Google Play indicated that over 45,000 mHealth apps are currently available. Surprisingly, despite the plethora of apps available to assist people in reducing problematic alcohol and other drug use, only a very small proportion of these apps are evidence based [14]. Evaluations of other health-related apps, for example in the field of mental health, have produced positive [15] or negligible [16] changes in the targeted behavior. Importantly, a number of early trials of apps that focused on problematic substance use have produced promising results, suggesting that apps could play a role in assisting individuals who are dependent on tobacco, alcohol, or illicit drugs to quit or maintain abstinence [17].

It has been approximately 10 years since the emergence of mobile apps designed to help people reduce or recover from problematic alcohol, tobacco, and/or illicit drug use [18]; 6 years since controlled trials have appeared [17]; and 5 years since the first publication of a systematic literature review of 6 smartphone apps for problematic substance use [19]. Of note, all but 1 review has examined the literature on digital interventions more broadly, and all reviews have only included alcohol interventions within their review. For example, Kaner et al [20] found that digital interventions (ie, delivered via a computer, smartphone, or mobile device) for alcohol use showed they significantly lowered alcohol consumption, with an average reduction of up to 3 (United Kingdom) standard drinks per week compared with control participants. Similarly, a meta-analysis showed that internet-delivered alcohol interventions significantly reduced problematic drinking behavior among adults, reducing by 5 standard units of alcohol consumption each week compared with the control group [21]. Berman et al [22] examined the use of mobile interventions (IVR, SMS, and apps) to reduce drinking in university students. A total of 2 of the 7 reviewed studies used apps, and they found that only an IVR intervention resulted in a reduction in the primary outcome. Finally, the most recent related review focused on alcohol use in community participants and similarly included a range of mobile interventions beyond apps [23]. Moreover, 5 of the 19 studies in their review [23] included app-based interventions with mixed findings reported. To date, there has been only 1 app-specific systematic review, which included pilot studies and open trials [19], which evaluated the alcohol app studies across community and alcohol-dependent individuals. The authors found that 2 of the 6 mobile apps reviewed reported reliable positive outcomes, with a further 2 showing promise. The authors highlighted the limited number of studies, small sample sizes, lack of control groups, and limited rigorous designs within the field of mobile app interventions for problematic substance use. Since their review, which was conducted in 2015, there has been a three-fold increase in controlled evaluations of mobile apps designed to reduce substance use or aid recovery from substance dependence. Although some reviews have included apps when examining the effectiveness of alcohol interventions delivered via mobile devices [20,22,23], surprisingly, there has been no further synthesis of the evidence regarding the effectiveness of problematic alcohol use interventions delivered specifically via mobile apps. Moreover, none of the app-specific reviews included smoking and illicit drug use to develop a comprehensive *picture* of the effectiveness of apps across the substance use field. Thus, this paper reports on the current evidence base regarding the effectiveness and feasibility of mobile apps designed to reduce problematic alcohol, tobacco, and illicit drug use. In addition, we report usability, adherence, retention, and engagement data where possible. This information is critical to gain a deeper understanding of user experience and behavior alongside effectiveness data. Thus, it has been approximately 10 years since the first appearance of an app targeting the reduction of substances, and it is timely for us to review the progress of the field.

Methods

Aims and Guidelines

A systematic search of the literature was performed to synthesize the findings on effects, retention, and usability from primary studies evaluating mobile app interventions to reduce tobacco, drug, and/or alcohol use. An initial scan of the literature indicated that there was a wide range of mobile delivery methods, considerable variability in intervention content, and control groups; hence, we decided a priori to not conduct a traditional meta-analytic review given the potential risk of drawing premature conclusions. The review focused on primary consumption outcomes (ie, quantity and/or frequency of

substance consumption) rather than related harm or secondary psychosocial outcomes. The search followed the PRISMA (preferred reporting items for systematic review and meta-analyses) guidelines [24]. Data extraction was guided by the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials-Electronic Health Checklist) [25]. Risk of bias was assessed using the *Risk of Bias* tool developed by the Cochrane Collaboration [26].

Literature Search and Screening

The literature search utilized the following large databases: MEDLINE (Medical Literature Analysis and Retrieval System Online), PsycInfo, EMBASE (Excerpta Medica dataBASE, via the OVID platform), and ERIC (Education Resources Information Center, via EBSCO). The databases were searched using variations of 3 key terms: Substance AND Intervention AND Smartphone App (refer to Multimedia Appendix 1 for the detailed search strategy).

Eligibility Criteria

The search was limited to papers published in peer-reviewed journals from January 1, 2007 (the year when the first smartphone was released) to February 1, 2019. Papers were eligible if they adhered to the following criteria: (1) they reported primary empirical data; (2) the primary focus of the intervention was reduction in the use of illicit drugs, and/or alcohol, and/or tobacco; (3) substance consumption outcome data were reported; (4) the intervention was delivered via a mobile device app (not web-based or SMS in isolation); and (5) a controlled trial design was employed using either a randomized or matched control methodology. There were no limitations on the type of control conditions employed. In addition, no language restrictions were imposed.

The software program Covidence (Veritas Health International) [27] was utilized to ensure independence of screening and accurate calculation of agreements. As shown in Figure 1, the combined search identified a total of 2714 potentially relevant articles (4 papers sourced from examining the reference lists of existing reviews) that reduced to 2100 after duplicates were removed. In keeping with the methodology proposed by Moskowitz [28] and Foxcroft et al [29], an abstract and title search was conducted independently on all papers by 2 people (KG and RO) with a random 20% of these cross-checked (by PS). Any disagreements were discussed by the authors (PS, RO, and PL) and an agreement made, resulting in a total of 31 potentially relevant articles. These 31 papers were read independently and in full by RO and PS. During this process, 14 additional articles were excluded. Agreement regarding exclusions was high (84%), with only 2 disagreements, which were then discussed with PL to reach a final decision. Each of the reference lists of the remaining papers were scanned, and papers known to the authors were included, which identified 3 additional studies, bringing the total number of studies informing this review to 20.

Figure 1. Flow diagram illustrating the selection process for the systematic review of the literature.
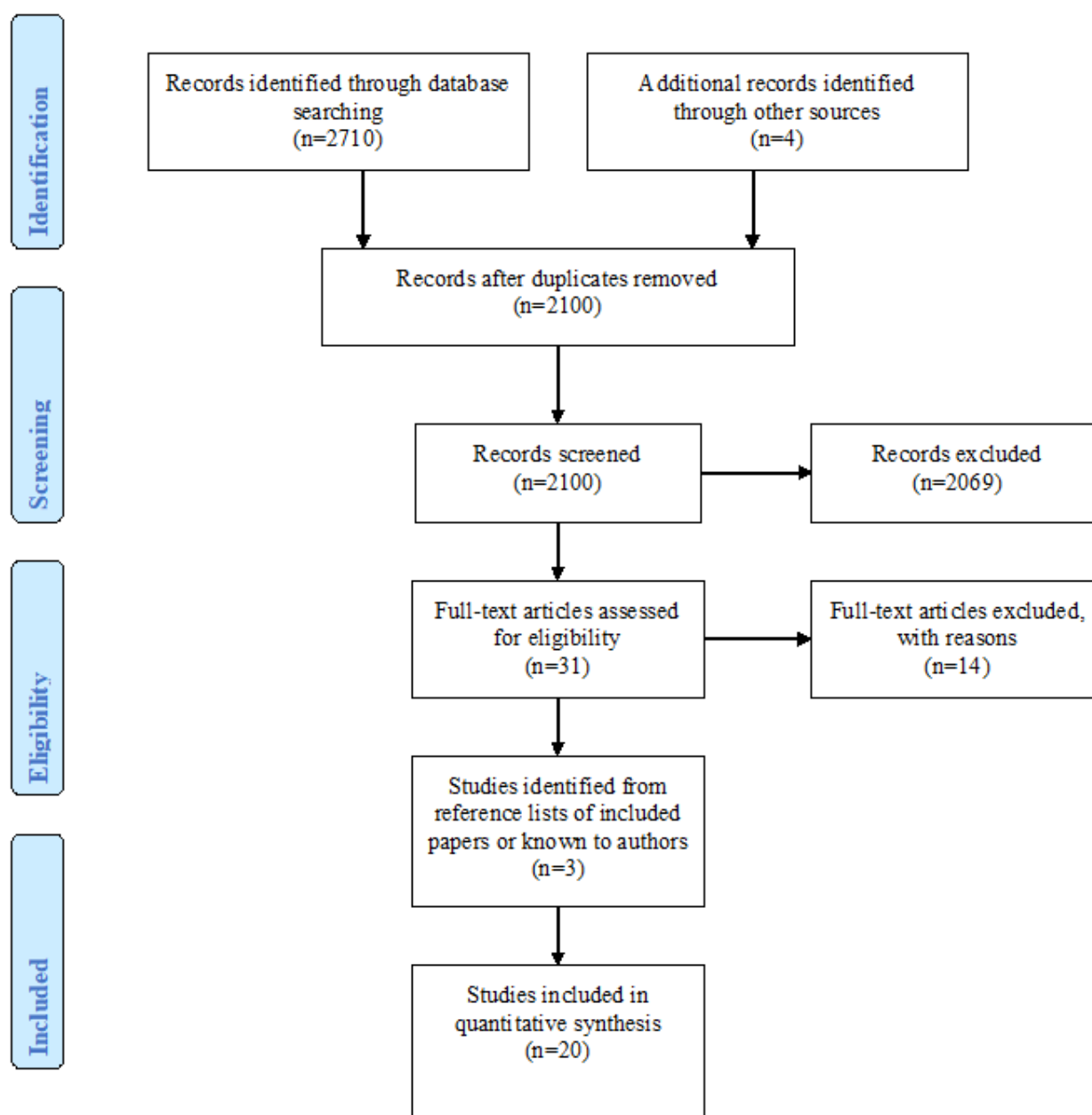
Data Extraction

The following data were extracted (RB and RO) from the 20 eligible articles: year, authors, study sample, target substance, consumption measure, length of intervention, description of the intervention and control and associated sample sizes, assessment times, summary of statistical evidence, effect sizes where possible (Tables 1 and 2). As indicated earlier, the only outcome data extracted were consumption variables, given that it is the most consistently reported variable, thus enabling comparability between studies. In this respect, means and SDs for all consumption outcomes for each group at baseline, post intervention, and follow-up (if reported) were extracted as well as the relevant statistical data and effect sizes (if reported). Some information was gained via contacting authors, and this is noted in Table 2. Retention and usability data were also extracted to provide an informed discussion regarding the feasibility of delivering interventions for substance use via an app (Table 3). Considerable data are reported in these tables, and hence, only brief summaries are reported in the text.

Table 1. Summary of studies.

Author, date [reference]; "app name"	Sample type; target substances	Intervention groups description	Comparison groups description	Intervention duration; assessment time points[a]; retention[b]	Age; gender
Aharonovich et al, 2017 [30]; "Health-Call" [c]	HIV-positive adults, drug and alcohol use during past month; drugs and alcohol	"HealthCall" consists of (a) Self-monitoring questions: primary drug; other drugs; drinking; HIV medication adherence; safe sex practices; wellness, stress. (b) Personalized feedback: graphs with goal attainment and feedback. (c) Option to call counselor. App use: daily notifications for 2 months. Adjunct components: (a) Motivational interviewing: face-to-face session at baseline, and two booster sessions.	Motivational interviewing as described for intervention group. Comparison includes app? No	Duration: 60 days; assessments: baseline, 60 days; retention: 91%	Whole sample: mean 50.96 (SD 7.04) years; 23.40% female
Baskerville et al, 2018 [31]; "Crush the Crave"	Young adults who smoked cigarettes daily and were considering quitting smoking in the next 30 days; tobacco	"Crush the Crave": (a) graphic performance feedback, (b) evidence-based information on relapse, craving management, (c) notifications on rewards, app use reminders; (d) access to quit lines, and (e) use of nicotine replacement therapy. App use: At will, with prompts to use the app. Adjunct components: "Crush the Craving" Facebook community; app delivered support and inspirational photos tailored to quit plan and stage of quitting; recording smoking so they understand triggers.	Printed self-help guide, "On the Road to Quitting"; (a) health benefits of quitting, (b) rewards of quitting, (c) smoking triggers, (d) coping with withdrawal and cravings, (e) setting a quit date, (f) seeking counseling and use of nicotine replacement therapy, (g) information on social support, (h) telephone support, (i) prevention of weight gain, and (j) relapse Comparison includes app? No	Duration: 6 months; assessments: baseline, 3 months, 6 months; retention: 61%	Intervention: aged range 19-29 years; 44.9% female; comparison: age range 19-29 years; 47.0% female
Boendermaker et al, 2015 [32]; "Alcohol/Avoid"	University students reporting regular drinking; alcohol	"Alcohol/Avoid" (a) cognitive bias modification using alcohol images; (b) participants swipe alcohol images away, and soft drink images toward themselves. App use: 14 days of access. Adjunct components: none.	Desktop computer version of Alcohol/Avoid training. Comparison includes app? No	Duration: 14 days; assessments: baseline, 28 days; retention: 81%	Whole sample: mean 22.44 (SD 2.58) years; 60.32% female
Bricker et al, 2014 [33]; "SmartQuit"	Adults who smoke at least five cigarettes daily; tobacco	"SmartQuit" consists of Acceptance Commitment Therapy and: (a) motivation and planning to quit, (b) acceptance of urges, and (c) self-compassion for slips. App use: 8-weeks access, no notifications, weekly emails. Adjunct components: none.	Use of app "QuitGuide" from National Cancer Institute. Similar features but no acceptance or self-compassion components: (a) reasons and plans to quit, (b) coping with slips and urges, and (c) staying positive. Comparison includes app? Yes.	Duration: 8 weeks; assessments: baseline, 2 months; retention: 82%	Intervention: mean 41.5 (SD 12.0) years; 53% female; comparison: mean 41.6 (SD 13.9) years; 51% female
Crane et al, 2018 [34]; "Drink Less"	Adults with AUDIT[d]>8, risky drinkers; alcohol	"Drink Less"–enhanced version: (a) goal setting, (b) personalized normative feedback, (c) cognitive bias retraining, (d) self-monitoring and feedback, (e) action planning, and (f) identity change.[e] App use: daily notifications to report consumption for 4 weeks, optional use of intervention modules. Adjunct components: none.	"Drink Less" minimal version consists of (a) goal setting; (b) alcohol psychoeducation; (c) sham cognitive bias retraining; (d) consumption self-monitoring; (e) information only about action planning; (f) information only on role of identity. Comparison includes app? Yes.[f]	Duration: 4 weeks; assessments: baseline, 4 weeks; retention: 27%	Both conditions: mean 39.2 (SD 10.9) years; 56% female

XSL•FO

RenderX

Author, date [reference]; "app name"	Sample type; target substances	Intervention groups description	Comparison groups description	Intervention duration; assessment time points[a]; retention[b]	Age; gender
Crane et al, 2018 [35]; "Smoke Free"	Smokers aged over 18 years who had set a quit date (no consumption criteria)	"Smoke Free" full version (a) goal setting and (b) delivery of messages that report on the benefits achieved since cessation attempt, iand "Daily Missions," which included behavior change techniques to resist cravings. App use: Daily messages delivered for 30 days from quit date. Adjunct components: none.	"Smoke Free" reduced version (a) goal setting; (b) self-monitoring; (c) delivery of messages of the benefits achieved since cessation attempt. Comparison includes app? Yes.	Duration: 12 weeks; assessments: baseline, 12 weeks; retention: 8%	Intervention: mean 28.7 (SD 9.0) years; 49% female; comparison: mean 29.1 (SD 9.4) years; 49% female
Davies et al, 2017 [36]; "Drinks Meter"	Adults aged 18-30 years who self-identified as a current drinker; alcohol	"Drinks Meter" consists of (a) Personalized feedback compared to individuals with similar demographics about alcohol use, calories consumed, and money spent. (b) Assessment (AUDIT) and brief advice and strategies regarding reduction. App use: At will, not stated whether prompts were sent. Adjunct components: none.	Comparison 1: website consists of (a) 20 questions about embarrassing events experienced while drinking and (b) tailored feedback on consequences of drinking. Comparison 2: control website consists of (a) being asked to imagine receiving information about alcohol. Comparison 3: assessment only. Comparison includes app. No.	Duration: 4 weeks; assessments: baseline, 4 weeks; retention: 82%	Whole sample: mean 21.70 (SD 3.28) years; 67.20% female
Earle et al, 2018 [37]; "CampusGANDR"	University students (no consumption criteria); alcohol	App intervention 1: "CampusGANDR" (PNF[+]) uses normative feedback and peer judgement. The game is played weekly with peers, whereby participants answer one alcohol-related and one nonalcohol-related question about their behavior. After 4 days, they receive normative feedback (ie, how their responses compared with their peers) and reflective evaluations from other students (ie, how they were judged by their peers). App use: questions and feedback delivered once each per week over 6 weeks. Adjunct components: None. App intervention 2: PNF—Same as app intervention above, with only normative feedback to alcohol questions (reflective evaluations were for nonalcohol-related questions).	Same as app intervention 1 condition, however, with no normative or reflective feedback on alcohol questions (normative feedback and reflective evaluations were given for nonalcohol-related questions in this condition). Comparison includes app? Yes	Duration: 6 weeks; assessments: baseline, 2 months; retention: 80% (PNF[+]); 84% (PNF)	Whole sample: age not provided; 55% female
Gajecki et al, 2014 [38]; "PartyPlanner"	University students reporting AUDIT>8 (men) or AUDIT>6 (women), risky drinkers; alcohol	PP[g] app consists of (a) real-time feedback on eBACs[h]; (b) simulating a drinking event by entering predicted eBAC levels before an event; (c) the user records their alcohol consumption then compares the simulation with real-life event; (d) tracks how drinking compares with safe drinking. App use: no notifications, instruction to use before drinking events. Adjunct components: none. PK[i] self-monitoring of alcohol; real-time eBAC feedback and alcohol-reduction strategies.	Assessment only comparison. Comparison includes app? No	Duration: 6 weeks; assessments: baseline, 7 weeks; retention: 61% (PP) and 74% (PK)	Whole sample: mean 24.72 (SD 4.81) years; 51.7% female

XSL•FO
RenderX

Author, date [reference]; "app name"	Sample type; target substances	Intervention groups description	Comparison groups description	Intervention duration; assessment time points[a]; retention[b]	Age; gender
Gajecki et al, 2017 [39]; "TeleCoach"	University students reporting excessive alcohol consumption (>9 drinks per week for women; >14 for men); alcohol	"TeleCoach" (a) reporting of alcohol consumption for a week; (b) brief feedback and psycho-education; (c) a relapse prevention skills training, and guided relaxation and mindful "urge-surfing." App use: no notifications, instructed to use at will. Adjunct components: eBAC app providing real-time feedback for 6 weeks before the intervention, with access during intervention.	Assessment only. Comparison includes app? No. However, as in the intervention condition, participants had access to another eBAC app for 6 weeks before, with access during intervention.	Duration: 12 weeks; assessments: baseline, 12 weeks; retention: 76%	Whole sample: mean 25.41 (SD 6.45) years; 69.1% female
Gonzalez and Dulin, 2015 [40]; "LBMI-A"	Adults who met the Diagnostic and Statistical Manual of Mental Disorders-5 alcohol dependence criteria; alcohol	"LBMI-A": (a) assessment and feedback, (b) high-risk locations for drinking, (c) using supportive people for change, (d) managing cravings, (e) problem-solving skills, (f) communication or drink refusal skills, and (g) pleasurable nondrinking activities. App use: feedback reports delivered weekly, daily interviews to report alcohol consumption. Adjunct components: none	"Drinker's Check-Up plus Bibliotherapy", a web-based intervention that includes assessment of drinking. objective and normative feedback, decisional balance exercise, goal selection, development of change plan, and links to other web-based resources. Comparison includes app? No.	Duration: 6 weeks; assessments: baseline, 6 weeks; retention: 60%	Intervention: mean 33.57 (SD 6.54) years; 46.4% female; comparison: mean 34.30 (SD 6.22) years; 35.0% female
Gustafson et al, 2014 [17]; "A-CHESS"	Individuals who met the DSM-IV[j] alcohol dependence criteria; alcohol	A-CHESS consists of (a) access to counselors, (b) a panic button related to relapse, (c) meditation, (d) recovery stories, (e) meeting locations, (f) recovery information, and podcasts App use: no notifications, instructed to use at will. Adjunct components: residential treatment.	TAU[k] (support offered through the residential service); comparison includes app: No	Duration: 8 months; assessments: 4 months (during intervention, no baseline); 8 months; 12 months; retention: 78%	Intervention: mean 38.3 (SD 9.5) years; 39.4% female; comparison: mean 38.4 (SD 11.2) years; 39.1% female
Hasin et al, 2014 [41]; "HealthCall"	Adults who were HIV-positive and alcohol dependent; alcohol	"HealthCall" (a) alcohol self-monitoring; (b) consumption feedback compared with drinking goal, and feedback on drinks per drinking day and reasons for drinking. App use: one prompt per day re self-monitoring. Adjunct components: counselors administered a 25-min motivational interviewing session.	Same as intervention, but app replaced by HealthCall-IVR (a daily phone call using interactive voice response). Comparison includes app? No	Duration: 60 days; assessments: baseline, 60 days; retention: 90%	Intervention: mean 45.5 (SD 11.5) years; 28.2% female; comparison: mean 46 (SD 7.2) years; 18.6% female
Hertzberg et al, 2013 [42]; "mCM"	Adults with post-traumatic stress disorder who were regular smokers; tobacco	"mCM" (a) using a CO[l] device to check abstinence; (b) using camera to record CO reading; (c) financial reward for each uploaded video showing "abstinent" CO, with progressive reinforcement schedule. App use: twice daily notifications for 4 weeks. Adjunct components: (a) two smoking cessation counseling sessions; (b) nicotine replacement therapy, low-nicotine cigarettes, and Bupropion; (c) 6 calls to assist withmotivation; (d) an additional 2 weeks of mCM app use, but without financial compensation.	Same app as intervention, but using noncontingent compensation, based on submitting videos of CO monitoring process, regardless of positive or negative CO reading. Comparison includes app? Yes	Duration: 4 weeks; assessments: 4 weeks, 3 months; retention: not reported	Intervention: mean 42.5 (SD 4.5) years; 36.4% female; comparison: mean 53.3 (SD 11.6) years; 27.3% female

Author, date [reference]; "app name"	Sample type; target substances	Intervention groups description	Comparison groups description	Intervention duration; assessment time points[a]; retention[b]	Age; gender
Hides et al, 2018 [43]; "Ray's Night Out"	People aged 16-25 years who drank alcohol at least monthly; alcohol	"Ray's Night Out" (a) information on harm minimization strategies, (b) motivation to set a drinking goal, and (c) psychoeducation on consequences of intoxication. App use: no notifications used. Adjunct components: none.	Comparison procedure: waitlist. comparison includes app: No.	Duration: 1 month; assessments: baseline, 1 month; retention: 96%	Intervention: mean 20.4 (SD 2.2) years; 79.2% female; comparison: mean 20.5 (SD 2.5) years; 76.2% female
Kerst and Waters, 2014 [44]; "AR Training"	Adults who smoked 10 or more cigarettes per day for the past 2 years; tobacco	"AR Training" consists of attentional retraining (cognitive bias modification). Each training involves 160 trials. Trial begins with fixation cross on screen, followed by picture pair (smoking and neutral image), then dot probe. Required to indicate dot probe location quickly. App use: Four daily notifications (one assessment and three training) for 7 days. Adjunct components: none.	Same as app intervention, except dot probe equally likely to replace smoking and neutral images (no attentional bias modification). Comparison includes app? Yes.	Duration: 7 days; assessments: baseline, day 8; retention: 94%	Intervention: mean 41.8 (SD 10.2) years; 50% female; comparison: mean 43.6 (SD 14.0) years; 47% female
Krishnan et al, 2018 [45]; "Coach2Quit"	Daily smokers aged 18-years and above.	"Coach2Quit" uses real-time data from a carbon monoxide exhaler to provide users with tailored messages based on their CO result which is also graphically displayed. App use: twice daily notifications following CO breath test. Adjunct components: brief advice; c	Comparison procedure: brief advice only. Comparison includes app? No.	Duration: 30 days; assessments: baseline, day 14, day 30; retention: 87%	Intervention: Median 53 years; 59% female; comparison: Median 51 years; 58% female
Liang et al, 2018 [46]; "S-Health"	Adults from methadone maintenance clinics with heroin or other substances use in the past 30 days; drugs and alcohol	"S-Health" consists of daily surveys designed to serve as both assessment and intervention. Users respond to questions about (a) cravings; (b) affect; (c) trigger thoughts, places, and situations; (d) responses to triggers; and (e) social context. App use: daily notifications to complete surveys. Adjunct components: daily educational text message.	Daily educational text message (information about HIV prevention and other educational materials). Comparison includes app? No.	Duration: 4 weeks; assessments: 1 week, 2 weeks, 3 weeks, 4 weeks (no baseline); retention: 98%	Intervention: mean 41.7 (SD 8.7) years; 64% female; comparison: mean 41.3 (SD 6.8) years; 83% female
Ruscio et al, 2016 [47]; "Brief-MP"	Adults who smoked 10 or more cigarettes per day for the past 2 years; tobacco	"Brief-MP" consists of five audio-guided mindfulness sessions on (a) "urge-surfing" the craving, (b) mindfulness of the breath, body, thoughts, and emotions. Five daily assessments probed craving, mindfulness, and affect. App use: asked to meditate once per day. Four random daily assessment notifications and one following meditation session. Adjunct components: none.	Same as intervention, except sham-meditation recordings (eg, nonjudgmental awareness replaced with self-evaluation). Comparison includes app? Yes.	Duration: 2 weeks; assessments: baseline, 2 weeks; retention: 75%	Intervention: mean 45.34 (SD 11.84) years; 50% female; control: mean 44.16 (SD 13.64) years; 55% female

Author, date [reference]; "app name"	Sample type; target substances	Intervention groups description	Comparison groups description	Intervention duration; assessment time points[a]; retention[b]	Age; gender
Witkiewitz et al, 2014 [48]; "BASICS-Mobile"	College students who engaged in at least one episode of heavy drinking during the past 2 weeks; alcohol and tobacco	"BASICS-Mobile" (a) daily monitoring, (b) normative feedback, (c) general or health information about drinking and smoking, (d) protective behavioral strategies, (e) alternative activities, (f) urge-surfing, and (g) decisional balance exercise. App use: 3 alerts per day for 14 days. Adjunct components: None. Comparison app: daily self-monitoring of alcohol consumption	Completed only initial screening, baseline assessment, and 1-month follow-up. Comparison includes app: No	Duration: 14 days; assessments: baseline, 1 month; retention: 94%	Whole sample: mean 20.5 (SD 1.7) years; 28% female

[a]Assessment time points reported here do not include assessments *during* the intervention.

[b]Retention is indicated only for the intervention group(s), defined as percentage completion of final (post intervention or follow-up) assessments.

[c]Studies in italics reported significant outcomes for intervention app at post intervention and/or follow-up compared with control. All studies delivered apps via smartphones, except for the study by Kerst and Waters [44], which used personal digital assistants (PDAs).

[d]AUDIT: Alcohol Use Disorders Identification Test.

[e]Factorial design; participants used either the enhanced or the minimal version of each component.

[f]The *Drink Less* app used a factorial randomized controlled trial design, whereby participants were randomized to 1 of the 32 groups, each receiving a different combination of intervention and comparison modules.

[g]PP: PartyPlanner.

[h]eBAC: estimated blood alcohol concentration.

[i]PK: Promillekoll.

[j]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

[k]TAU: treatment as usual.

[l]CO: carbon monoxide.

Table 2. Summary of effects.

Study [reference]; target substance and substance use outcome measures	Intervention			Control			Between groups statistic and significance	Effect size (d)[a,b]	Quality assessment[c]
	N	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)	n	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)			
Aharonovich et al, 2017 [30]; drugs and alcohol[d]									
Number days primary drug use over 30 days	21	12.8 (4.4)	5.0 (4.7)	21	15.3 (7.3)	8.5 (6.1)	IRR[e]=.59 (.35-.99), P=.04	.17	Good
Number drinking days over previous 30 days	21	14.2 (7.3)	7.0 (7.6)	21	13.7 (6.3)	8.1 (5.7)	IRR=.67 (.41-1.07), P=.09	.23	Good
Standard drinks (14 g)[f] per day over previous 30 days	21	3.2 (2.4)	.9 (1.0)	21	2.7 (2.0)	1.3 (1.1)	IRR=.63 (.36-1.11), P=.11	.41	Good
Baskerville et al, 2018 [31]; tobacco									
Continuous abstinence at 6 months	354	0	49 (13.8)	371	0	57 (15.4)	OR[g]=.89 (.59-1.34), P=.56	.06	Good
7-day point prevalence abstinence at 6 months	342	0	114 (33.3)	366	0	143 (39.1)	OR=.78 (.57-1.06), P=.11	.14	Good
30-day point prevalence abstinence at 6 months	344	0	84 (24.4)	365	0	107 (29.3)	OR=.78 (.56-1.09), P=.14	.14	Good
Boendermaker et al, 2015 [32]; alcohol[h]									
Standard drinks (10 g) over 14 days	25	25.1 (21.4)	20.0 (17.1)	24	25.4 (19.1)	18.4 (14.6)	$F_{1,44}$=0.03, P>.05	−.05	Poor
Bricker et al, 2014 [33]; tobacco									
Percent of sample abstinent over 30 days	80	(0) 0%	(10) 13%	84	0%	(7) 8%	OR=2.7 (.8-10.3), P=.12	.55	Fair
Crane et al, 2018 [35]; tobacco									
Self-reported abstinence at 12-weeks	ITT[i]: 14,228; PP[j]: 1213	(0) 0%; (0) 0%	(234) 1.6%; (234) 19.3%	ITT: 13,884; PP: 901	(0) 0%; (0) 0%	(124) 0.9%; (124) 13.8%	OR=1.86 (1.49-2.31), P<.001; OR=1.50 (1.18-1.90), P<.001	.34; .22	Good
Crane et al, 2018 [34]; alcohol									
Average number of alcohol units (8 g) per week	NF[k]:336; CB[l]: 336; MF[m]: 336; AP[n]: 336; IC[o]: 336; NFxCB[p]: 168	39.1 (25.97); 40.3 (28.23); 39.9 (27.09); 39.0 (26.46); 39.0 (26.62); 38.9 (26.92)	24.5 (22.45); 27.2 (25.96); 26.3 (23.41); 23.8 (24.23); 27.1 (26.47); 23.2 (20.68)	336; 336; 336; 336; 336; 168	40.7 (28.66); 39.6 (26.45); 39.9 (27.63); 40.9 (28.20); 40.8 (28.05); 39.9 (27.85)	26.6 (26.60); 23.9 (22.79); 24.5 (25.56); 27.0 (24.53); 23.7 (21.82); 21.5 (21.09)	$F_{1,178}$=.30, P=.59; $F_{1,178}$=.39, P=.53; $F_{1,178}$=.78, P=.38; $F_{1,178}$=.13, P=.71; $F_{1,178}$=2.16, P=.14; $F_{4,43}$=4.68, P=.03	.08; .09; .13; .05; .22; .67	Good

Study [reference]; target substance and substance use outcome measures	Intervention			Control			Between groups statistic and significance	Effect size (d)[a,b]	Quality assessment[c]
	N	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)	n	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)			
Davies et al, 2017 [36]; alcohol									
AUDIT-C[q] (alcohol consumption score)	104	6.6 (2.62)	6.0 (2.33)	OTM[r]: 99; IC[s]: 97; WL[t]: 102	6.2 (2.54); 6.8 (2.49); 6.6 (2.46)	5.7 (2.47); 6.1 (2.38); 5.9 (2.42)	All between-group comparisons P=ns	−.01; .03; .03	Good
Earle et al, 2018 [37]; alcohol									
Maximum drink (undefined) number on single night over current semester	PNF+[u]: 72; PNF[v]: 79	4.23 (4.14); 3.87 (4.07)	2.97 (3.25); 3.53 (3.38)	71	3.80 (4.09)	3.82 (4.28)	See footnote$	.31; .09	Good
Drink (undefined) number over previous weekend	PNF+: 72; PNF: 79	3.08 (4.10); 2.65 (3.74)	1.94 (2.67); 2.26 (3.28)	71	3.32 (4.81)	3.21 (4.70)	See footnote$	.23; .07	Good
Composite score (maximum single occasion and drink number on previous weekend)	PNF+: 72; PNF: 79	3.08 (4.10); 2.65 (3.74)	1.94 (2.67); 2.26 (3.28)	71	3.32 (4.81)	3.21 (4.70)	Mean=−0.14, SE=0.07, P<.01[w]; mean=0.00, SE=0.07, P=ns[w]	.46; .24	Good
Kerst and Waters, 2014 [44]; tobacco									
Percent of sample any smoking on test day	30	(29) 97%	(27) 90%	30	100%	(27) 90%	B=.04 (−.35 to .43), P=.85	−.01	Fair
Expired carbon monoxide (ppm)	30	15.9 (5.35)	15.5 (7.70)	30	15.6 (4.68)	14.2 (5.81)	B=1.07 (−4.13 to 1.99), P=.49.	−.20	Fair
Salivary cotinine (ng/ml)	30	394 (181)	410 (211)	30	420 (253)	408 (268)	B=−23.4 (−105.4 to 58.6), P=.57	−.13	Fair
Gajecki et al, 2014 [38]; alcohol									
Standard drinks (12 g) over 7 days	PP[x]: 153; PK[y]: 341	8.57 (6.12); 9.62 (6.26)	8.32 (6.45); 9.75 (7.05)	489	9.15 (6.18)	8.62 (6.28)	Time x group LMM[z]P=.82; time x group LMM P=.41	−.05; −.11	Fair
Number drinking days over 7 days	PP: 153; PK: 341	2.17 (1.12); 2.24 (1.20)	2.17 (1.23); 2.36 (1.23)	489	2.29 (1.19)	2.15 (1.19)	Time x group LMM P=.23; time x group LMM Z=3.39, P=.001	−.12; −.22	Fair
Gajecki et al, 2017 [39]; alcohol[aa]									
Standard drinks (12 g) over 7 days	71	16.58 (7.84)	12.87 (9.73)	124	17.16 (7.87)	14.52 (7.46)	Z=−1.07, P=.29	.15	Fair
Number drinking days over 7 days	71	3.35 (1.20)	2.51 (1.15)	124	3.53 (1.39)	3.02 (1.45)	Z=−2.12, P=.03	.30	Fair
Gonzalez and Dulin, 2015 [40]; alcohol[aa]									

Study [reference]; target substance and substance use outcome measures	Intervention			Control			Between groups statistic and significance	Effect size (d)[a,b]	Quality assessment[c]
	N	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)	n	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)			
Standard drinks (14 g) over 7 days	28	39.12 (20.37)	22.07 (22.08)	20	33.34 (21.58)	12.26 (13.19)	PE=−4.28, $P<.05$	See footnote[ab]	Poor
Percent heavy drinking days[ac]	28	54.25 (28.93)	26.98 (29.47)	20	47.74 (29.71)	15.04 (24.03)	PE=−8.25, $P<.05$	See footnote[ab]	Poor
Percent abstinent days	28	30.36 (22.48)	51.32 (32.25)	20	38.21 (29.85)	60.90 (39.22)	PE=10.29, $P<.05$	See footnote[ab]	Poor
Gustafson et al, 2014 [17]; alcohol									
Number of risky drinking days over 30 days[ad]	132	1.50 (0.34)[ae]	Post: 1.54 (0.49); FU[af]: 1.13 (0.50)	139	3.01 (0.48)[ae]	Post: 2.65 (0.48); FU: 2.60 (0.49)	Post mean difference=1.11, $P=.10$; FU mean difference=1.47, $P=.03$	−.18; .24	Poor
Percent of sample abstinent over 30 days	132	(100) 75.6%[ae]	Post: (103) 78.1%; FU: (104) 78.7%	139	(94) 67.7%[ae]	Post: (93) 66.9%; FU: (91) 65.5%	Post between group OR=1.70, $P=.04$; FU between group OR=1.94, $P=.02$	.29; .37	Poor
Hasin et al, 2014 [41]; alcohol									
Standard drinks (14 g) per drinking day over 30 days	39	9.3 (6.9)	3.9 (2.1)	43	8.1 (3.9)	3.6 (1.6)	See footnote[ag]	.16	Poor
Percent of abstinent days over 30 days	39	58.1 (27.4)	79.2 (22.5)	43	61.3 (24.2)	82.1 (17.8)	See footnote[ag]	.01	Poor
Hertzberg et al, 2013 [42]; tobacco									
Percent of sample abstinent for previous 7 days (bio-verified) at end of 4-week treatment	11	(0) 0%	(9) 82%	11	(0) 0%	(5) 45%	$X^2_1=1.6$, $P=.21$	.55	Poor
Percent self-report (not bio-verified) abstinence at 3-months post intervention	11	(0) 0%	(6) 55%	11	0%	(2) 18%	$X^2_1=2.0$, $P=.15$	.64	Poor
Hides et al, 2018 [43]; alcohol									
Risky single occasion drinking frequency over 1 month[ah]	97	2.11 (0.91)	2.23 (1.17)	86	2.10 (1.08)	2.25 (1.15)	$F_{1,184}=0.28$, $P=.60$	.03	Fair
Typical standard drinks (10 g) over 1 month	97	2.79 (1.41)	2.56 (1.32)	86	2.64 (1.40)	2.24 (1.21)	$F_{1,188}=1.00$, $P=.32$	-.12	Fair
Krishnan et al, 2018 [45]; tobacco									

XSL·FO
RenderX

Study [reference]; target substance and substance use outcome measures	Intervention			Control			Between groups statistic and significance	Effect size (d)[a,b]	Quality assessment[c]
	N	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)	n	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)			
Percent self-reported and biochemically verified abstinence at 30 days	39	(0) 0%	(1) 3%	50	0%	(1) 2%	$P>.05$	.04	Poor
Median number of cigarettes over 30 days (IQR)	39	10.0 (6.0-20.0)	5 (4.0-10.0)	50	10.0 (8.0-20.0)	6.0 (3.0-10.0)	$P=.84$	See footnote[aj]	Poor
Median expired carbon monoxide in ppm (IQR)	39	23.0 (18.0-33.0)	19.5 (15.0-26.0)	50	22 (14.0-30.0)	18.5 (10.0-28.0)	$P=.52$	See footnote[aj]	Poor
Liang et al, 2018 [46]; drugs									
Number of days with primary drug use over 7 days (pre=end of week 1, no baseline)	48	1.33 (2.48)	0.71 (1.87)	25	3.08 (3.37)	2.20 (3.06)	OR=.29 (.06-1.44), $P=.13$ (regressed over 4 time points during intervention)	.09	Poor
Drugs detected in urine (pre=end of week 1, no baseline)	42	(23) 56%	(11) 26%	20	(14) 70%	(10) 50%	OR=.57 (.11-2.84), $P=.49$ (regressed over 4 time points during intervention)	.11	Poor
Ruscio et al, 2016 [47]; tobacco									
Cigarettes per smoking day	20	N/A	N/A	17	N/A	N/A	$F_{1,436}=5.50$, $P=.01$[ak]	N/A	Fair
Expired carbon monoxide (ppm)	18	18.4 (9.8)	14.3 (8.2)	14	19.1 (6.7)	15.4 (5.0)	$F_{1,29}=0.01$, $P=.92$	.05	Fair
Salivary cotinine (ng/ml)	18	504.4 (300.3)	433.9 (257.1)	14	452.9 (221.9)	482.8 (250.0)	$F_{1,29}=0.04$, $P=.84$	.37	Fair
Witkiewitz et al, 2014 [48]; alcohol and tobacco									

Study [reference]; target substance and substance use outcome measures	Intervention			Control			Between groups statistic and significance	Effect size (d)[a,b]	Quality assessment[c]
	N	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)	n	Pre, mean (SD) or n (%)	Post, mean (SD) or n (%)			
Standard drinks (14 g) per drinking day	BM[al]: 30; DM[am]: 29	BM: 5.57 (2.81); DM: 5.58 (2.45)	BM: 4.83 (2.59); DM: 4.56 (2.65)	26	7.46 (3.46)	6.05 (2.88)	$WALDX^2_2=0.1$, $P=.94$ [an]	.08	Poor
Number of heavy drinking days over 7 days[ao]	BM: 30; DM: 29	BM: 2.31 (1.53); DM: 2.45 (1.44)	BM: 2.07 (1.70); DM: 1.76 (1.33)	26	2.86 (1.41)	2.31 (1.35)	$WALDX^2_2=0.1$, $P=.91$[an]	.07	Poor
Cigarettes per smoking day	BM: 30; DM: 29	BM: 4.93 (3.43); DM: 4.78 (4.83)	BM: 3.28 (3.35); DM: 2.71 (2.86)	26	3.76 (2.15)	4.55 (4.07)	B=2.04, P=.002; B=1.59, P=.02	.55[ap]; .45[ap]	Poor
Number of days with drinking and smoking over 7 days	BM: 30; DM: 29	BM: 2.81 (0.59); DM: 2.82 (0.64)	BM: 1.97 (1.09); DM: 2.07 (0.88)	26	2.66 (0.48)	1.76 (0.83)	$WALDX^2_2=2.0$, $P=.38$ [an]	.31	Poor

[a]All effect sizes are Cohen d. Sign of effect size indicates agreement with hypothesized direction (positive implies app condition improved outcome to a greater degree than comparison conditions; ie, a reduction in consumption or an increase in rates of abstinence).

[b]Where effect sizes not reported as Cohen's d, effect sizes were converted from reported effect sizes where possible or derived using pooled baseline SDs from intervention and control groups, as described by Morris [49].

[c]Good quality=all criteria in the Cochrane Risk of Bias tool were met, fair quality=one criterion not met or two criteria unclear and the assessment that this was unlikely to have biased the outcome and there was no important limitation that could invalidate the results, poor quality=one criterion not met or two criteria unclear and the assessment that this likely biased the outcome and there were important limitations that could invalidate the results OR two or more criteria listed as high or unclear risk of bias.

[d]Studies in italics reported significant outcomes for intervention app at post-intervention and/or follow-up timepoints compared with control. Sample sizes reflect the number of participants included in the final analyses.

[e]IRR: incidence rate ratio.

[f]Amount in grams of pure alcohol in one standard drink varies across countries and is indicated in brackets.

[g]OR: odds ratio.

[h]Some data provided directly from authors.

[i]ITT: intention to treat analysis (referred to in publication as "Missing Equals Smoking").

[j]PP: per protocol analysis (referred to in publication as "Follow-up Only").

[k]NF: personalized normative feedback.

[l]CB: cognitive bias retraining.

[m]MF: monitoring and feedback.

[n]AP: action planning.

[o]IC: identity change.

[p]Two-way interaction between personalized normative feedback and cognitive bias retraining.

[q]First three questions of the 10-item Alcohol Use Disorder Identification Test, probing alcohol consumption.

[r]OTM: "One Too Many" (intervention website).

[s]IC: imagery control (sham website).

[t]wl: waitlist control.

[u]Intervention using personalized normative feedback plus reflection

[v]Intervention using personalized normative feedback without reflection.

[w]Main analysis used variables derived by combining drinking measures and controlling for a range of other covariates.

[x]PP: PartyPlanner app.

[y]PK: Promillekoll app.

[z]LMM: linear mixed model.

[aa]Some data provided directly from authors.

[ab]Group×time interaction analysis during intervention (not pre-post and hence was not considered superior to control as per our definition).

[ac]Heavy drinking defined here as 4+ standard drinks for females and 5+ for males.

[ad]Risky drinking defined here as 3+ standard drinks (14 g of alcohol) for females and 4+ for males consumed within a 2-hour period.

[ae]No baseline data were collected in this study as participants were inpatients who had not consumed alcohol for some time; authors use 4-month data as reference for 8-month post intervention and 12-month follow-up analyses.

[af]FU: follow-up.

[ag] "No between groups significance conducted.

[ah]Risky single occasion drinking in this study is defined as 5+ standard drinks (10 g of alcohol) during one occasion.

[aj]Median scores, cannot compute Cohen d effect size.

[ak]LMM group×day interaction based on daily smoking reports over 2 weeks (not pre-post).

[al]BM: BASICS-Mobile app.

[am]DM: Daily monitoring app.

[an]Omnibus chi-square test across all 3 conditions.

[ao]Heavy drinking defined here as 4+ standard drinks for females and 5+ for males.

[ap]Effect size controlling for range of predictors.

[$] Between group significance testing not conducted

Table 3. Summary of usability.

Study [reference]; app name	Usability measures	Usability
Aharonovich et al, 2017 [30]; HealthCall	Engagement—proportion of days used out of total possible days, and satisfaction—7 items rated 1 (low) to 5 (high).	Engagement: 95% (range 68.7%-100%) of the possible 60 days. Satisfaction: mean 4.5 (SD 0.8).
Baskerville et al, 2018 [31]; Crush the Crave	Four satisfaction items (used frequently, easy to use, well laid out, and confidence in using) measured on a 5-point Likert scale ("strongly agree" to "strongly disagree"), overall satisfaction item using same scale as above, an overall helpfulness item on a 10-point Likert scale.	Mean (SD): Used frequently: 3.6 (1.2); Easy to use: 2.3 (1.1); Well laid out: 2.5 (1.1); Confidence in using: 2.8 (1.1); Overall satisfaction: 2.6 (1.3); Overall helpfulness; 4.3 (2.7).
Boendermaker et al, 2016 [32]; Alco-hol/Avoid	User experience measured on 5-point Likert scale, 1 (strongly disagree) to 5 (strongly agree).	Mean (SD): Ease of use:15.77 (2.11); Player enjoyment:13.19 (2.98); Player involvement Mean (SD) 11.23 (2.13); Task compliance:6.07 (1.53).
Bricker et al, 2014 [33]; SmartQuit	Treatment satisfaction measured on 5-point scale, 1 (not at all) to 5 (very much), and utilization: self-reported number of times opened app.	85% said app organized; 53% said app useful for quitting; 59% were satisfied overall; app mean use 37 times over 8 weeks (no prompts).
Crane et al, 2018 [34]; Drink Less	Four usability measures rated on a 5-point scale, 1 (not at all) to 5 (extremely).	Mean (SD): Helpfulness: NF[a]: 3.05 (0.88); CB[b]: 3.02 (0.98); MF[c]: 3.18 (0.93); AP[d]: 3.04 (1.02); IC[e]:3.09 (0.97). Ease of use: NF: 3.45 (0.97); CB: 3.45 (0.97); MF: 3.59 (1.00); AP: 3.56 (1.07); IC: 3.57 (1.00). Recommend: NF: 2.99 (1.23); CB: 2.91 (1.23); MF: 3.25 (1.22); AP: 3.08 (1.23); IC: 3.15 (1.16) . Satisfaction: NF: 3.22 (0.95); CB: 3.20 (0.97); MF: 3.36 (1.00); AP: 3.26 (1.00); IC: 3.25 (0.95).
Dulin et al, 2014 [50]; LBMI-A	Helpfulness and ease of use of each tool rated on a 7-point scale, 1 (extremely unhelpful or extremely easy to use) to 7 (extremely helpful or extremely difficult to use).	Mean (SD): Ease of use: 5.6 (1.7), with the Drink Monitor Tool being the easiest to use, 6.6 (1.8) and the High-Risk Location Tool being the most difficult, 4.3 (1.6). Helpfulness: High-Risk Location Tool was least helpful, 3.8 (2.2) and the Daily Interview Tool was most helpful, 6.1 (1.1).
Gajecki et al, 2014 [38]; PartyPlanner	Self-reported app usage and questions on ease of use, suitability, and likelihood of recommending to a friend, on a 5-point scale.	Mean (SD): Self-reported usage (any): PP[f], 41.4%; PK[g], 74.1%. Ease of use: PP: 3.2 (1.1); PK: 4.0 (1.1). Suitability: PP: 3.6 (1.2); PK 3.4 (1.2). Would recommend: PP: 3.6 (1.2); PK 3.7 (1.3).
Hasin et al, 2014 [41]; HealthCall	Twelve satisfaction questions on a 3-point scale, assessing feelings of safety and privacy, effects on recall and knowledge of own drinking patterns, motivation and self-confidence to reduce consumption, and app's ability to prompt drinking goals.	Intervention group ("agree"): perceived responses as safe (94.59%), concerned about privacy (37.84%), liked using app (91.89%), graphs increased interest in app (86.49%), and graphs increased perceived benefit of app (91.89%). Of the 30 daily suggestions for cutting down drinking, 13 were rated as "helpful"/"very helpful" by over half the patients.
Hides et al, 2018 [43]; Ray's Night Out	Mobile Application Rating Scale (5-point rating scale, 23 items), assessing engagement, functionality, aesthetics, and information quality.	Mean (SD): The MARS[h] indicated the app had a good level of overall app quality: 3.82 (0.51); Functionality: 3.98 (0.69); Aesthetics: 4.03 (0.62); Information: 4.0 (0.56). Participants reported that they were unlikely to pay for the app:1.25, (0.69) and gave it a 3 out of 5-star rating: 3.13 (0.76).
Liang et al, 2018 [46]; S-Health	Seven usability questions (5-point scale), assessing ease of use, recall feasibility, willingness to provide responses, etc.	Intervention group ("strongly agree" or "agree"): "The survey questions were easy to understand" (55.3%); "I was comfortable answering these questions" (68.1%); "I was able to remember the number of days or frequency using alcohol or drugs in the past week" (53.2%); "The smartphone screen was easy to use" (72.3%); "I prefer to answer these questions myself on a cellphone instead of having a person ask me" (46.8%).
McTavish et al, 2012 [51]; A-CHESS	Passive app use data: which service selected; duration of use for each service; which pages viewed; messages sent or received.	93.5% accessed the system during the first week after leaving treatment. The A-CHESS services used by the greatest percentage of participants included *discussions*, *my messages*, *my team*, and *weekly surveys*. The least used services were *instant library*, *frequentlyasked questions*, and *web links*.
Pocuca et al, 2016 [52]; Ray's Night Out	Mobile Application Rating Scale - youth version (5-point rating scale, 23 items), assessing engagement, functionality, aesthetics, and information quality.	Mean (SD): Entertaining: 3.78 (SD 0.83); Interesting: 3.67 (0.71); Customizable: 3.00 (0.58); Interactive: 2.63 (0.74); Speed and accuracy of function: 4.78 (0.44); Ease of use: 4.44 (0.73); Flow and logic: 4.33 (0.50); Layout design: 4.33 (0.50); graphics quality 4.56 (0.73); quality and credible information 3.71 (0.58); would you recommend 2.78 (1.09); overall star rating 3.11 (0.60).

XSL·FO
RenderX

[a]NF: personalized normative feedback.

[b]CB: cognitive bias retraining.

[c]MF: monitoring and feedback.

[d]AP: action planning.

[e]IC: identity change.

[f]PP: PartyPlanner app.

[h]MARS: Mobile App Rating Scale.

[g]PK: Promillekoll app.

Risk of Bias

Risk of bias was assessed using the *Risk of Bias* tool developed by the Cochrane Collaboration [26]. This tool enabled the assessment of 5 key sources of bias: (1) allocation sequence, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, and (5) incomplete outcome data. The assessment of risk of bias for each study was conducted by 2 independent reviewers (JB and RB), and disagreement was resolved through discussion. Table 2 provides the quality ratings of the included studies; please refer to Figure 2 [17,30-48] for a summary of the sources of bias.

Figure 2. Risk of bias summary.

Data Synthesis and Analyses

As stated, a decision was made a priori to not conduct a meta-analysis. Nonetheless, we report Cohen *d* effect sizes to enable meaningful comparisons and interpretations (Table 2). When effect sizes were not reported in the paper, we computed them using pooled baseline SDs from the intervention and control groups, as described by Morris [49].

Results

Design and Target Sample

A total of 20 studies met the inclusion criteria for the review. Eighteen studies were randomized controlled trials (RCTs) and two were matched controlled studies [39,41], as the control groups in each of these studies were from a related trial

(conducted by the same research group), matched on main measures, and adopting the same eligibility criteria (ie, matched controlled studies). Specifically, in the study by Gajecki [39], participants in the comparison were from an assessment-only control group in a concurrent study that had the same eligibility criteria and were matched on alcohol consumption. In the study by Hasin [41], the matched controls were the control participants from a previous randomized trial adopting the same eligibility criteria.

The studies varied considerably in the sample size per group available for analysis, ranging from 22 to 14,228. The majority of studies were recruited from community samples with a wide age range, with the exception of 5 studies that included only university students. Five studies included a clinical or dependent sample [17,40-42,46], and gender tended to be evenly balanced, with the exception of the clinical studies, which had a greater proportion of males. Moreover, 10 apps targeted alcohol reduction[17,32,34,36,38-41,43,45], 7 targeted tobacco [31,33,35,42,44,47], 1 targeted alcohol and tobacco [48], 1 targeted drugs [46], and 1 targeted illicit drugs and alcohol [30].

Risk of Bias

The methodological quality varied considerably. A total of 11 studies (11/20, 55%) reported adequate generation of random sequencing, 9 studies (9/20, 45%) reported adequately concealed group allocation, 5 studies (5/20, 25%) reported appropriate blinding of participants and personnel, 13 studies (13/20, 65%) reported appropriate blinding of outcome assessment, and 14 studies (14/20, 70%) adequately accounted for incomplete data. Overall, only 3 studies were classified as having a low risk of bias on all 5 measures of risk of bias. Moreover, 4 studies were classified as having a low risk of bias on 4 measures, 5 studies on 3 measures, 1 study on 2 measures, and 7 studies had low risk of bias on either none or only one measure. See Figure 2; more details regarding the source of bias and an in-text description of the risk of bias are provided in Multimedia Appendix 1.

App Content, Complexity, and Supportive Components

Apps differed substantially in their intervention content. Most apps were stand alone, but 8 [17,30,31,39,41,42,45,46] of them had additional adjunct components such as supportive counseling; motivational interviewing; educational messages; links to resources; peer group supports such as Facebook groups; nicotine replacement therapy; audio-guided relaxation; and even a high-risk patient locator, which sends an alert to patients if they are approaching a high-risk drinking location (Table 1). In terms of app complexity, at least half of the apps consisted of multiple intervention elements [17,34,35,48]; others were simple and employed a single distinct intervention such as approach-bias training [44,47]. Apps varied substantially in the underlying theoretical approaches of their interventions (motivational interviewing, approach-bias modification, meditation, acceptance and commitment therapy, and relapse prevention). Finally, the majority of apps included self-monitoring of substance use as part of the intervention, and only 1 app (DrinkLess) directly tested the intervention components.

Comparison Conditions

The comparison conditions were highly diverse. That is, Crane et al [34], Crane et al [35], Earle et al [37], Hertzberg et al [42], Kerst and Waters [44], and Ruscio et al [47] used a variant or minimal version of the intervention app, controlling for some key intervention component; Bricker et al [33] used an unrelated app; Boendermaker et al [32] and Hasin et al [41] used the same intervention but not delivered by an app; Gonzalez and Dulin [40], Liang et al [46], and Davies et al [36] used a nonapp different substance use intervention; Aharonovich et al [30] and Gustafson et al [17] used the adjunct or treatment as usual intervention as the comparison group (ie, minus the app); Baskerville et al [31] used information material and links to resources; and 5 studies used a waitlist, passive control or assessment only [36,38,39,43,48].

Intervention Duration, Time to Follow-Up, Notifications, and Frequency of Contact

The intervention duration was generally short to medium length (1 to 8 weeks), with the exception of Gajecki et al (12 weeks) [39], Hides et al (26 weeks) [43], and Gustafson et al (35 weeks) [17]. *Smoke Free* and *Drink Less* did not specify an intervention length. Aside from 2 studies [17,42], no follow-up assessments were conducted following end-of-treatment measures. Most apps employed assessment and/or intervention notifications or alerts; however, 6 apps did not use notifications and instead requested that participants use the app at will or during specific events such as when drinking alcohol [17,32,33,38,39,43]. For the apps that did employ notifications, the most common schedule was once per day. Four apps used more frequent notifications 2 times per day [42], 3 times per day [48], and 4 times per day [44,47]. One app employed a single weekly email reminder [33].

Effectiveness Outcomes

Cohen *d* effect sizes were extracted or calculated for each substance consumption outcome. For five studies where outcome data did not conform to the requirements for the calculation as described by Morris [49], effect sizes were computed by converting from the reported effect size to *Cohen d,* as indicated in Table 2. In three cases, insufficient data were provided (despite requests to authors) to calculate a pre-post effect size.

A minority of studies (6/20; Table 2) reported significant reductions in substance use compared with the comparison group at post treatment or follow-up.

Of the 6 apps that reported superior outcomes (compared to controls) at post treatment, 3 targeted alcohol, one of which was with clinical participants (A-CHESS), and the other two were focused on university students (TeleCoach and CampusGANDR). One app was delivered to smokers (SmokeFree) and another targeted both alcohol and smoking in a university population (BASICS-Mobile), but in the latter study, only the smoking reductions (not alcohol) were superior to the comparison condition. Finally, one app (HealthCall) targeted illicit drugs and alcohol in an HIV population, but only a significant reduction in drug use (not alcohol) was reported in comparison to controls. It should be noted that most apps did report reductions in substance use;

however, they were not necessarily superior to the control conditions post treatment. In addition, three apps were categorized as showing promise. Brief-MP (smoking app) and LBMI-A (alcohol app with a clinical group) reported intervention effects during treatment but not post treatment, and DrinkLess reported a significant reduction in alcohol consumption with a combination of two components (normative feedback and cognitive training).

Witkiewitz et al [48] reported on "BASICS-Mobile" that delivered monitoring, normative feedback, health information, alternative activities, and "urge-surfing" over 14 days focusing on alcohol and smoking reduction in university students. The app performed better for cigarettes smoked per day compared with a minimal control condition at post intervention (d=0.55), and no intervention effect was found for alcohol. Gustafson et al [17] showed that "A-CHESS," which delivered psychoeducation, recovery stories, meeting locations, guided meditation, and access to phone counselors over 8-months, performed better than treatment as usual for 30-day alcohol abstinence (d=0.37) and number of risky drinking days (d=0.24). Aharonovich et al [30] showed that, alongside motivational interviewing, "HealthCall"—an app delivered over 8 weeks employing motivational self-monitoring, personalized feedback, and the option to call a phone counselor—produced significantly lower rates of primary drug use compared with a motivational interviewing only control group (d=0.17). Gajecki et al [39] reported that "TeleCoach"—an app that delivered alcohol monitoring, personalized feedback, alcohol guidelines and risk situations, drink refusal skills, and "urge-surfing" over 3 months—was associated with a reduction in drinking occasions (but not quantity) compared with a waitlist group (d=0.30). Earle et al [37] reported that the app "CampusGANDR"—a campus-based game that primarily centered on normative and injunctive feedback over 6 weeks—was associated with a reduction in drink number over a weekend compared with a control app that provided feedback about activity reports unrelated to drinking (d=0.23). Crane et al [35] reported that "Smoke Free"—an app consisting of goals, monitoring, daily messages that reported on accrued benefits (eg, financial savings and estimated health improvements), and behavior change strategies over a 30 day period—was associated with higher 3-month continuous smoking abstinence rates compared with a minimal version of the app (d=0.22 using per protocol analysis and d=0.34 using intention to treat analysis; see Multimedia Appendix 1 for details).

In addition, we categorized three further apps as showing promise. Brief-MP and LBMI-A were associated with significant reductions in the intervention arm, although this difference was no longer significant at post intervention (see Multimedia Appendix 1 for details) [40,47]. In addition, although *Drink Less* was associated with no overall difference between the intervention and control app, a significant interaction was found between the normative feedback and cognitive training components within the intervention group only, suggesting that these two components in combination resulted in a greater decrease in alcohol consumption compared with their minimal app (see Multimedia Appendix 1 for details). Given that this

analysis was exploratory (although prespecified), we await further research before drawing any conclusions.

Retention, Engagement, and Usability

Retention rates for the intervention group—usually defined as completion of final assessments (post or follow-up)—were generally good (70-80%) [17,39,47], high (80-90%) [32,33,37,42], or very high (over 90%) [30,41,43,44,46,48]. Lower retention rates were found in the studies by Gajecki et al (61%) [38], Gonzalez et al (60%) [40], and Crane et al (27%) [34]. See Table 3 for details.

Engagement—generally defined as responding to notifications or use in line with instructions (eg, in some trials, participants were instructed to use the app daily without providing notifications)—varied more, with some studies reporting low engagement (below 50%) [38,46], others reporting moderate engagement (50-80%) [47,51], and some reporting high user engagement (80% and over) [30,33,41,42,44].

Studies have used various methods to ascertain usability and user satisfaction, including reliable instruments such as the mobile application rating scale (MARS) [53], or a single item (Table 3). Satisfaction ratings ranged from moderate (50-80%) [33] to high (80% and over) [30,41]. For example, Hides et al [43] used the MARS and found that Ray's Night Out had good objective app quality and high (80% and over) levels of functionality, aesthetics, and information. Hasin et al [41] reported high satisfaction, with 86% of patients stating that HealthCall-S reminded them of their drinking goal and over 80% stating that it increased confidence and motivation to reduce drinking. In the study by Bricker et al [33], 59% said they were satisfied overall.

In summary, of the 6 apps that were significantly more effective than their comparison conditions, all reported small to moderate effect sizes. Moreover, 3 of the 6 app studies were assessed as having a high risk of bias and 3 as having a low risk of bias; hence, no particular pattern emerged regarding outcomes and bias. When multiple substance consumption measures were reported, significant outcomes were mostly variable. Further details of each study are provided in Multimedia Appendix 1 and the tables.

Discussion

The primary aim of this paper was to synthesize and report on an up-to-date systematic literature review focused on the effectiveness of substance use (alcohol, illicit drugs, or tobacco) interventions delivered via mobile apps. A total of 20 studies were included in the review, of which only 6 reported significantly greater reductions in substance use post intervention compared with comparison groups [17,30,35,37,39,48]. The average effect sizes were modest, although this is consistent with mobile apps in other fields, including mental health [15] and diet and exercise [54]. Two further trials [40,47] reported significant intervention effects during the treatment phase, with no significant group differences at post intervention. A third app reported a significant interaction for two intervention components within the app [34].

The 6 apps that performed significantly better than their comparison conditions varied substantially in intervention length, content, and complexity, and few commonalities across the majority of these emerged. In terms of app content, 3 of the 6 apps included normative feedback, and 1 app included personalized feedback (actual consumption compared with goals). Specifically, CampusGANDR rested heavily on personalized normative feedback and injunctive feedback (what peers think you should do); TeleCoach provided personalized normative feedback immediately following consumption reports; BASICS-mobile delivered normative feedback every day; and HealthCall included personalized feedback comparing actual consumption with personal goals. Interestingly, in the earlier study by Gajecki [38], the comparison condition, which included personalized normative feedback, performed better than the intervention, which did not deliver normative consumption feedback. This association between personalized feedback and normative feedback is consistent with previous face-to-face interventions demonstrating the effectiveness of these approaches [55] within mHealth approaches to substance reduction. Given the known importance of peers and normative attitudes in relation to substance use, including this component in future apps (particularly in young populations) may enhance efficacy.

In addition, the length of intervention may have played a role in influencing positive outcomes, with only 1 of the 6 highlighted interventions being under 6 weeks long (4, 6, 8, 12, and 35 weeks). In contrast, most of the studies that did not report intervention superiority ran for 4 or fewer weeks. Although only suggestive, it is possible that behavior change via an app may be more effective when the intervention component is greater than 4 weeks and participants engage for longer periods.

Retention rates were generally high across all studies, with the majority showing above 90% retention at postintervention or follow-up, except for the *Smoke Free* [35] and *Drink Less* [34] apps where retention at follow-up was less than 30%. The two latter studies differed from the rest as people enrolled in the study after having downloaded the app, as opposed to being recruited to a trial from the outset. This suggests that retention may be poor for apps used outside of research trials, and methods that enhance retention are of utmost importance if apps are to be effective as a public health approach. Moreover, 9 of the 20 studies reported usability data, with some variable results. Encouragingly, participants generally experience mHealth apps as easy and convenient to use. Considering that the poor usability of smartphone apps is common and can substantially compromise user engagement [56], these results are promising.

Four of the apps (LBMI-A, A-CHESS, Health Call, and Health S) were targeted at clinical samples who were primarily alcohol-dependent individuals, except for Health S (heroin addiction). Only 1 of the 4 reported superior outcomes (A-CHESS) [17]. This app intervention was for alcohol-dependent individuals who had already been in residential treatment, and hence, it functioned as a relapse prevention program. Furthermore, A-CHESS included adjunct components such as contact with counselors when required, indicating that the app alone was likely not responsible for the intervention effect. In addition, it is likely that this study has

some risk of bias as it would have been clear to participants that they were in the intervention group given that the comparison group was treatment as usual with no additional support. It is not surprising given the complexity of alcohol and drug dependence that a mobile app may not result in significant positive outcomes for clinical samples, particularly given that most of the interventions except for A-CHESS were 6 weeks or less in duration. Although it is too early to draw any firm conclusions, it does appear that if mobile apps are helpful for those who are dependent on substances, it is likely to be most effective as posttreatment support rather than as the primary intervention. Interestingly, of the 7 apps that targeted tobacco use (1 targeted both alcohol and tobacco, BASICS), the only 2 reporting superior outcomes were *Smoke Free* and BASICS, which had no lower limit on smoking level, whereas the other trials included daily smokers, some of which were smoking 10 cigarettes a day, which would be considered in the mild to moderate dependence range. Once again, this tends to suggest that individuals with heavier substance use (if not clinical) are less likely to benefit from mobile apps. Given the small numbers, it is not possible to draw any conclusions regarding effectiveness in relation to the type of substance, although there is some suggestion that mobile apps are less effective with dependent individuals. We await further studies to confirm this conclusion.

Finally, it is important to note that the reductions in substance use produced by some of the app interventions were small in absolute terms. For example, compared with the comparison conditions, the app conditions with significant consumption outcomes produced mean reductions of one less day of drug use over 30 days [30], 0.8 of a day less drinking per week [39], 5% increase in the likelihood of being abstinent [17], and one less drink over a weekend [37]. Nonetheless, at a public health level, even small reductions at a population level can have a significant impact on the reduction of mortality and morbidity associated with problematic alcohol and other drug use and thus remain encouraging. Furthermore, although the majority of studies did not report "superior" outcomes (to their comparison conditions), in many cases, they reported significant decreases in alcohol or illicit drugs or tobacco. This will occur in study designs when comparison conditions are other apps or interventions delivered via other digital modalities (web-based and IVR) that we also know have a positive impact on substance reduction. In this respect, the mHealth field would benefit from greater consensus and clarity regarding the expectations of app efficacy and the role mobile apps should play in clinical treatment and public health approaches.

Limitations

Numerous limitations were apparent in the included studies. For example, most studies were affected by design limitations and risk of bias and many were small sample pilot studies. That is, studies varied considerably in terms of sample size, with many small studies (eg, eight intervention conditions had samples of 30 or fewer); in contrast, two studies had samples greater than 1000.

One of the most significant limitations was that comparison conditions varied considerably, and in many cases were poorly

balanced with the intervention condition. This variability in design reflects distinct kinds of research questions and precludes being able to draw any conclusions (tentative or otherwise) about the effectiveness of apps compared with other modes of delivery for problematic substances. Similarly, the 6 apps reporting superior outcomes compared with controls are confounded by substantially different kinds of comparison conditions—some likely to have very little therapeutic benefit (eg, waitlist control) and others comprised a similarly comprehensive "treatment" as the intervention condition (eg, web-based version of the same intervention). In addition, some studies included comparison conditions that were poorly balanced in terms of content and frequency of contact [17,30,46]. For example, in one study [17], participants in the treatment group had more counselor contact and completed a weekly assessment of alcohol intake, not delivered in the comparison condition, which may have produced an assessment effect. The Bricker et al study [33] included unspecified adjunct therapies (intervention group participants were encouraged to use other therapies alongside the trial, with no reporting of the details).

Finally, the risk of bias was generally high. At times, this was due to lack of detailed information, so it is possible that the true risk of bias could be lower across the studies. Overall, only 6 studies were classified as having either no or low risk of bias (Table 2; Multimedia Appendix 1). A further seven studies were assessed as potentially being biased, but a lack of information did not enable them to be classified as low risk. The remaining 8 studies were assessed as having a high risk of bias. Moreover, 3 of the 6 studies reporting superior outcomes were assessed as having no risk or low risk of bias, providing confidence that half of the significant findings were highly robust. Two of the superior trials had some risk of bias, of which one was due to unclear descriptions and one had high risk.

Future Recommendations

This review highlights a number of key areas for future work in this fast-growing area. First, it is clear that we need sufficiently powered trials with longer follow-up periods and greater attention to reduce the potential risk of bias in these studies. An increasing focus on protocol papers, pre-registered trials, and adherence to Cochrane guidelines (and reporting thereof) will result in the ability to draw stronger conclusions in the next review.

Second, this review highlighted considerable variability in app content and complexity across a range of substances and inadequate descriptions of app content within publications. Furthermore, only half of the studies included descriptions of the user experience, which is critical to consider alongside the effectiveness data. If engagement and satisfaction from the user perspective is low, then the effectiveness outside of trial studies will be very low. The lack of usability data can be partly explained by word constraints and the reluctance of some journals to publish "user experience" papers. Thus, we recommend the field to engage with the Open Sciences Framework and similar platforms when providing details of app content, theories of change, and design.

In some cases, the development of app interventions was clearly described within the context of a theory of change for substance use reduction (eg, within the papers reporting on Smoke Free, Drink Less, and BASICS-mobile). However, in many cases, it was unclear what the proposed mechanism of change was and why it was chosen, and at times, it was difficult to ascertain the content of the intervention. In some cases, the rationale was to transfer "effective" face-to-face treatments to mobile apps (ie, BASICS-mobile), whereas other authors developed bespoke app interventions dependent on user input and the unique aspects of smartphone technology (ie, Ray's Night Out). Ultimately, despite the substance use field having now produced 20 controlled evaluations of mobile apps, we remain unclear as to which "types" of interventions are likely to be most effective and the theory of change model underpinning them. Finally, in most cases, except for the DrinkLess app, there was little investigation of the effectiveness of the intervention components. The positive interaction between cognitive bias training and normative feedback found in the post-hoc analyses of the DrinkLess app is promising, given the ease by which both of these intervention components can be translated into a mobile app. Furthermore, the cognitive training component is a habit-forming activity and is well suited to an intervention that can be easily attended to on a smartphone at any time. Importantly, some behavior change interventions may be more aligned to digital delivery than others. For instance, we have recently proposed that a time-based goal setting technique rather than traditional count-based goals (to reduce smoking or alcohol or drug use) could be substantially enhanced by the unique capabilities of app functionality [57]. This might include daily reminders, timed alerts, automating reduction goals, supportive psychological strategies, and personalized delivery of interventions. As we continue to make further technological advancements in app delivery, well-aligned intervention content will be critical to the success of mHealth. A fine-grained analysis of the content of mobile apps in this field would be a helpful exercise in future publications.

Third, it was surprising to see a lack of iterative co-design processes being described in many of the publications (although Smoke Free, Drink Less, and Ray's Night Out were some of the exceptions). Although potentially omitted in some cases due to manuscript length constraints, usability testing before evaluating the app in larger RCTs is critical. Such usability testing allows researchers to then modify the functionality based on user and clinician feedback, thereby avoiding inefficient or highly limited RCTs. Greater emphasis on co-design and usability testing will enhance our ability to improve retention. For example, we know that therapist guidance reduces attrition in digital interventions, but this can be costly. One possibility would be to trial automated guides or coaches to provide support and reduce attrition. Furthermore, the use of personalized reminders and strategies, machine-learning functionality, passive-sensor reporting, and context-based reminders have the potential to increase retention, in addition to other uses. However, none of the apps in this review incorporated these more complex and sophisticated technologies. We found this surprising, a pattern that could in part reflect funding constraints. Although evidence is lacking as to what level of collaboration is appropriate and at which point during the design process, it

is likely that greater interdisciplinary and co-design collaboration, including users, researchers, clinicians, software developers, policy makers, marketing teams, and graphic designers, can produce more sophisticated products that will leverage these capabilities in the context of university research trials.

Finally, the considerable range of comparison conditions was a major limitation of this review, with the rationale for some of the chosen comparison conditions being somewhat perplexing. Each different type of comparison condition reflects a different research question and implies quite different purposes for an app focused on substance reduction. Researchers could consider whether their trial seeks to determine if the tested app will produce superior effects to an identical, similar, or different app intervention; computer intervention; face-to-face intervention; treatment as usual; or no intervention. At a public health level, we propose that if an app reduces substance use to the same degree as a more costly intervention, then this should be considered a positive trial outcome. This was not a discussion engaged in most of these papers and does point to a broader policy-based discussion as to what constitutes "app effectiveness" to ensure transparent communication with the public. Such considerations are important, given the potential broad reach that apps have in remote and financially disadvantaged communities or in addressing numerous other barriers to help seeking. The current saturation of smartphones in our society makes them a powerful mHealth tool, but further work is needed to understand how best to harness their capabilities, engage the user, and generate positive intervention outcomes. With hopeful anticipation, we look forward to what the next 5 years in mHealth research and development brings.

Conclusions

It has been approximately 10 years since substance use interventions delivered via smartphone apps have become available, and the majority of controlled evaluations have been published in the last 5 years. As we are likely to see an acceleration in the development of smartphone app substance use interventions over the coming years, it is timely to take stock of the field and identify strengths, limitations, and future directions. This state-of-the-art review highlights the diversity in app design, with a range of options being explored for both community and clinical populations. The review also highlights substantial variability in study design, intervention types, comparison conditions, measures, follow-up period, length of

intervention, and reporting details, making it almost impossible to infer factors or themes associated with the effectiveness of substance use apps specifically. We see this review as a *taking stock* moment; we are clearly not at the point where any firm conclusions can be drawn. Importantly, guidance from the details and outcomes of this review will hopefully strengthen the mHealth field in its future endeavors to assist individuals in the community to reduce their problematic consumption of alcohol, tobacco, and/or illicit drugs. Ultimately, we hope that mHealth can provide affordable, accessible, and effective behavior change interventions in this field.

Co-design is critically important in all intervention development; however, many studies do not incorporate the user until after important design and intervention decisions are made. To answer whether an app intervention is equivalent or superior in efficacy to other formats, the app should be tested against the same intervention delivered within a nonapp comparison condition. Ultimately, comparison conditions should be selected based on the fundamental research question. In addition to app-specific functionality that can be leveraged to produce innovative interventions, apps that demonstrate at least outcome equivalence compared with face-to-face treatment or treatment as usual would offer numerous advantages, including low cost, accessibility, reduced barriers to help seeking, and potentially higher engagement. Relatedly, efficacious app interventions that are able to recruit individuals otherwise unwilling to seek help would also offer substantial advantages in addressing the treatment gap. Indeed, app interventions have generated considerable interest in public health research, with some promising signs emerging from mental health apps [15], although see study by Weisel et al [16]. However, a similar story cannot yet be told for apps focused on helping people reduce problematic alcohol, tobacco, and/or illicit drug use. Although the field is still in its infancy, this review cautiously suggests that app interventions for problematic substance use are yet to clearly demonstrate their utility. In particular, and not surprisingly, this seems to be the case for clinical or heavier users of substances. A more positive state of the literature in the next review is likely to be enabled by greater collaboration between multidisciplinary teams, iterative learning from each other's products, selecting evidence-based and mobile app–aligned content, greater expert and consumer input, attention to reducing risk of bias, comprehensive usability testing, more personalized interventions, and methods that leverage greater user engagement and retention.

Authors' Contributions

PS and RO conceptualized the review and screened the abstracts and titles of identified papers. RO and RB extracted the data. All authors, except for JM, contributed to writing the first draft of the manuscript. PL converted the data to *d* effect sizes where possible and independently checked data extraction. JM contributed to the interpretation of the results, future recommendations, and decisions regarding quality assessment. RB and JB conducted the quality assessment, which was independently checked by PS. All authors approved the final submission of the manuscript.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 22 KB - jmir_v22i11e17156_app1.docx]

References

1. Gowing LR, Ali RL, Allsop S, Marsden J, Turf EE, West R, et al. Global statistics on addictive behaviours: 2014 status report. Addiction 2015 Jun;110(6):904-919. [doi: 10.1111/add.12899] [Medline: 25963869]
2. Chen L, Strain EC, Crum RM, Mojtabai R. Gender differences in substance abuse treatment and barriers to care among persons with substance use disorders with and without comorbid major depression. J Addict Med 2013;7(5):325-334 [FREE Full text] [doi: 10.1097/ADM.0b013e31829b7afe] [Medline: 24091763]
3. Kimergård A, Foley M, Davey Z, Dunne J, Drummond C, Deluca P. Codeine use, dependence and help-seeking behaviour in the UK and Ireland: an online cross-sectional survey. QJM 2017 Sep 1;110(9):559-564. [doi: 10.1093/qjmed/hcx076] [Medline: 28379496]
4. Lubman DI, Cheetham A, Jorm AF, Berridge BJ, Wilson C, Blee F, et al. Australian adolescents' beliefs and help-seeking intentions towards peers experiencing symptoms of depression and alcohol misuse. BMC Public Health 2017 Aug 16;17(1):658 [FREE Full text] [doi: 10.1186/s12889-017-4655-3] [Medline: 28814325]
5. Wallhed Finn S, Bakshi A, Andréasson S. Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. Subst Use Misuse 2014 May;49(6):762-769. [doi: 10.3109/10826084.2014.891616] [Medline: 24601784]
6. Saunders SM, Zygowicz KM, D'Angelo BR. Person-related and treatment-related barriers to alcohol treatment. Journal of Substance Abuse Treatment 2006 Apr 01:261-270. [doi: 10.1016/j.jsat.2006.01.003]
7. Oleski J, Mota N, Cox BJ, Sareen J. Perceived need for care, help seeking, and perceived barriers to care for alcohol use disorders in a national sample. Psychiatric Services 2010 Dec 1:1223-1231. [doi: 10.1176/ps.2010.61.12.1223.]
8. Park-Lee E, Lipari RN, Hedden SL, Copello EAP, Kroutil LA. Receipt of services for substance use and mental health issues among adults: Results from the 2015 National Survey on Drug Use and Health. NSDUH Data Review 2016:e-e [FREE Full text] [Medline: 29431966]
9. Sawares AS, Shen N, Xue Y, Abi-Jaoude A, Wiljer D. The impact of mobile apps on alcohol use disorder: a systematic review protocol. JMIR Res Protoc 2017 Apr 4;6(4):e49 [FREE Full text] [doi: 10.2196/resprot.6975] [Medline: 28377366]
10. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]
11. Mhealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2011.
12. Smart Everything, Everywhere - Mobile Consumer Survey 2017: The Australian Cut. Deloitte. 2017. URL: https://www2. deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/ deloitte-au-tmt-mobile-consumer-survey-2017-211117.pdf [accessed 2020-10-08]
13. Lee H, Ahn H, Choi S, Choi W. The SAMS: smartphone addiction management system and verification. J Med Syst 2014 Jan;38(1):1. [doi: 10.1007/s10916-013-0001-1] [Medline: 24395031]
14. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
15. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
16. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
17. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May 1;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
18. Kim J, Ryu D, Park J, Kim H, Yu Y. Development of an Android Application for Sobriety Test Using Bluetooth Communication. In: International Conference on Multimedia, Computer Graphics, and Broadcasting. 2011 Dec 10 Presented at: MCGB'11; December 8-10, 2011; Jeju Island, Korea. [doi: 10.1007/978-3-642-27204-2]
19. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Advanced HealthCare Technologies 2015 Dec;1:47-54. [doi: 10.2147/ahct.s65791]

XSL•FO
RenderX

20. Kaner E, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]

21. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: an individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]

22. Berman AH, Gajecki M, Sinadinovic K, Andersson C. Mobile interventions targeting risky drinking among university students: a review. Curr Addict Rep 2016;3:166-174 [FREE Full text] [doi: 10.1007/s40429-016-0099-6] [Medline: 27226948]

23. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]

24. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

25. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

26. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]

27. Covidence Systematic Review Software. Veritas Health Innovation. URL: http://www.covidence.org/ [accessed 2020-09-07]

28. Moskowitz JM. The primary prevention of alcohol problems: a critical review of the research literature. J Stud Alcohol 1989 Jan;50(1):54-88. [doi: 10.15288/jsa.1989.50.54] [Medline: 2648075]

29. Foxcroft DR, Lister-Sharp D, Lowe G. Alcohol misuse prevention for young people: a systematic review reveals methodological concerns and lack of reliable evidence of effectiveness. Addiction 2006 Jan 24;92(5):531-537. [doi: 10.1111/j.1360-0443.1997.tb02911.x]

30. Aharonovich E, Stohl M, Cannizzaro D, Hasin D. HealthCall delivered via smartphone to reduce co-occurring drug and alcohol use in HIV-infected adults: a randomized pilot trial. J Subst Abuse Treat 2017 Dec;83:15-26 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.013] [Medline: 29129192]

31. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]

32. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]

33. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 1;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]

34. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]

35. Crane D, Ubhi HK, Brown J, West R. Relative effectiveness of a full versus reduced version of the 'smoke free' mobile application for smoking cessation: an exploratory randomised controlled trial. F1000Res 2018;7:1524 [FREE Full text] [doi: 10.12688/f1000research.16148.2] [Medline: 30728950]

36. Davies EL, Lonsdale AJ, Hennelly SE, Winstock AR, Foxcroft DR. Personalized digital interventions showed no impact on risky drinking in young adults: a pilot randomized controlled trial. Alcohol Alcohol 2017 Nov 1;52(6):671-676. [doi: 10.1093/alcalc/agx051] [Medline: 29016711]

37. Earle AM, LaBrie JW, Boyle SC, Smith D. In pursuit of a self-sustaining college alcohol intervention: deploying gamified PNF in the real world. Addict Behav 2018 May;80:71-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.005] [Medline: 29407688]

38. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Jul 2;9:11 [FREE Full text] [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]

39. Gajecki M, Andersson C, Rosendahl I, Sinadinovic K, Fredriksson M, Berman AH. Skills training via smartphone app for university students with excessive alcohol consumption: a randomized controlled trial. Int J Behav Med 2017 Oct;24(5):778-788 [FREE Full text] [doi: 10.1007/s12529-016-9629-9] [Medline: 28224445]

40. Gonzalez VM, Dulin PL. Comparison of a smartphone app for alcohol use disorders with an internet-based intervention plus bibliotherapy: a pilot study. J Consult Clin Psychol 2015 Apr;83(2):335-345 [FREE Full text] [doi: 10.1037/a0038620] [Medline: 25622202]

XSL•FO

RenderX

41. Hasin DS, Aharonovich E, Greenstein E. HealthCall for the smartphone: technology enhancement of brief intervention in HIV alcohol dependent patients. Addict Sci Clin Pract 2014 Feb 17;9:5 [FREE Full text] [doi: 10.1186/1940-0640-9-5] [Medline: 24533631]

42. Hertzberg JS, Carpenter VL, Kirby AC, Calhoun PS, Moore SD, Dennis MF, et al. Mobile contingency management as an adjunctive smoking cessation treatment for smokers with posttraumatic stress disorder. Nicotine Tob Res 2013 Nov;15(11):1934-1938 [FREE Full text] [doi: 10.1093/ntr/ntt060] [Medline: 23645606]

43. Hides L, Quinn C, Cockshaw W, Stoyanov S, Zelenko O, Johnson D, et al. Efficacy and outcomes of a mobile app targeting alcohol use in young people. Addict Behav 2018 Feb;77:89-95. [doi: 10.1016/j.addbeh.2017.09.020] [Medline: 28992580]

44. Kerst WF, Waters AJ. Attentional retraining administered in the field reduces smokers' attentional bias and craving. Health Psychol 2014 Oct;33(10):1232-1240. [doi: 10.1037/a0035708] [Medline: 24818609]

45. Krishnan N, Elf JL, Chon S, Golub JE. COach2Quit: a pilot randomized controlled trial of a personal carbon monoxide monitor for smoking cessation. Nicotine Tob Res 2019 Oct 26;21(11):1573-1577 [FREE Full text] [doi: 10.1093/ntr/nty182] [Medline: 30169740]

46. Liang D, Han H, Du J, Zhao M, Hser Y. A pilot study of a smartphone application supporting recovery from drug addiction. J Subst Abuse Treat 2018 May;88:51-58 [FREE Full text] [doi: 10.1016/j.jsat.2018.02.006] [Medline: 29606226]

47. Ruscio AC, Muench C, Brede E, Waters AJ. Effect of brief mindfulness practice on self-reported affect, craving, and smoking: a pilot randomized controlled trial using ecological momentary assessment. Nicotine Tob Res 2016 Jan;18(1):64-73. [doi: 10.1093/ntr/ntv074] [Medline: 25863520]

48. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychol Addict Behav 2014 Sep;28(3):639-650 [FREE Full text] [doi: 10.1037/a0034747] [Medline: 25000269]

49. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. Organ Res Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]

50. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]

51. McTavish FM, Chih M, Shah D, Gustafson DH. How patients recovering from alcoholism use a smartphone intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]

52. Pocuca N, Hides L, Zelenko O, Quek L, Stoyanov S, Tulloch K, et al. Initial prototype testing of Ray's Night Out: A new mobile app targeting risky drinking in young people. Computers in Human Behavior 2016 Jan;54:207-214. [doi: 10.1016/j.chb.2015.07.033]

53. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

54. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

55. Reid A. Social norms and health behavior. Handbook of behavioral medicine: Methods and applications 2019;Springer:263-274. [doi: 10.1007/978-0-387-09488-5_19]

56. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]

57. Staiger PK, Liknaitzky P. An alternative goal‐setting technique for addictive behaviour interventions: the Chronos approach. Clin Psychol 2020 May 23:-. [doi: 10.1111/cp.12226]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials-Electronic Health Checklist
IVR: interactive voice response
MARS: mobile application rating scale
mHealth: mobile health
RCT: randomized controlled trial

XSL•FO

RenderX

Adaptive Elements in Internet-Delivered Psychological Treatment Systems: Systematic Review

Suresh Kumar Mukhiya[1], MSc; Jo Dugstad Wake[2], PhD; Yavuz Inal[3], PhD; Ka I Pun[1], PhD; Yngve Lamo[1], Prof Dr

[1]Western Norway University of Applied Sciences, Bergen, Norway

[2]NORCE Norwegian Research Centre, Bergen, Norway

[3]University of Bergen, Bergen, Norway

Corresponding Author:
Suresh Kumar Mukhiya, MSc
Western Norway University of Applied Sciences
Inndalsveien 28
Bergen, 5063
Norway
Phone: 47 55 58 58 00
Email: skmu@hvl.no

Abstract

Background: Internet-delivered psychological treatments (IDPTs) are built on evidence-based psychological treatment models, such as cognitive behavioral therapy, and are adjusted for internet use. The use of internet technologies has the potential to increase access to evidence-based mental health services for a larger proportion of the population with the use of fewer resources. However, despite extensive evidence that internet interventions can be effective in the treatment of mental health disorders, user adherence to such internet intervention is suboptimal.

Objective: This review aimed to (1) inspect and identify the adaptive elements of IDPT for mental health disorders, (2) examine how system adaptation influences the efficacy of IDPT on mental health treatments, (3) identify the information architecture, adaptive dimensions, and strategies for implementing these interventions for mental illness, and (4) use the findings to create a conceptual framework that provides better user adherence and adaptiveness in IDPT for mental health issues.

Methods: The review followed the guidelines from Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The research databases Medline (PubMed), ACM Digital Library, PsycINFO, CINAHL, and Cochrane were searched for studies dating from January 2000 to January 2020. Based on predetermined selection criteria, data from eligible studies were analyzed.

Results: A total of 3341 studies were initially identified based on the inclusion criteria. Following a review of the title, abstract, and full text, 31 studies that fulfilled the inclusion criteria were selected, most of which described attempts to tailor interventions for mental health disorders. The most common adaptive elements were feedback messages to patients from therapists and intervention content. However, how these elements contribute to the efficacy of IDPT in mental health were not reported. The most common information architecture used by studies was tunnel-based, although a number of studies did not report the choice of information architecture used. Rule-based strategies were the most common adaptive strategies used by these studies. All of the studies were broadly grouped into two adaptive dimensions based on user preferences or using performance measures, such as psychometric tests.

Conclusions: Several studies suggest that adaptive IDPT has the potential to enhance intervention outcomes and increase user adherence. There is a lack of studies reporting design elements, adaptive elements, and adaptive strategies in IDPT systems. Hence, focused research on adaptive IDPT systems and clinical trials to assess their effectiveness are needed.

(J Med Internet Res 2020;22(11):e21066) doi:10.2196/21066

Introduction

Research accounts for internet-delivered psychological treatment (IDPT) as a useful therapeutic tool [1] with increased potential to provide evidence-based mental health interventions for the far-reaching population at a lower cost [2,3]. However, actual user adherence to such interventions is low [4-8]. These results raise a critical question in IDPT: how can the clinical effect of IDPT be improved? Therefore, it is relevant to focus on the factors associated with enhancing user adaption toward such interventions.

Some studies have found that providing therapist contact for online guidance and support during interventions increases adherence and effect sizes [7,9-14]. Clarke et al [15] added telephone calls and postcard reminders from therapists to increase user adherence. The study concluded by discovering no significant difference between intervention groups with or without reminders. However, a similar study done by Farrer et al [16] to evaluate the effectiveness of a 6-week IDPT for depression with and without telephone interaction concluded that IDPT is effective both with and without tracking for reducing depression. The results indicated that the success of an intervention depends on the environmental settings in which they are performed [16]. Similarly, some studies increased the frequency of emails from therapists to increase user adherence. Klein et al [17] conducted a study to examine if the frequency of therapists' contact (from 1 email per week to 3 emails per week) made a difference in user adherence. The study concluded that the effectiveness of IDPT might be independent of the frequency of therapist support. Hilvert-Bruce et al [18] conducted a study to investigate if the drop out of users of IDPT was caused by lack of treatment efficacy, if changes in the choice of treatments, reminders, and financial cost improves adherence, and finally, if the addition of clinical contact improves user adherence. The results of the study showed that adding reminders, and increasing the choice of treatments, cost, timing, and contact with clinicians improved user adherence [18]. These findings illustrate that the baseline predictors of adherence vary across studies.

A systematic review by Christensen et al [7] discovered that disease diversity, treatment length, and predicted chronicity are essential factors contributing to user adherence in IDPT. Similarly, clinical severity has also been indicated as one of the crucial factors contributing to user adherence in web-based interventions targeting problematic drinking [19]. Similar factors have been identified as the most prominent factors in user adherence toward IDPT. However, only a few studies have discussed why the target group chose not to adhere to any specific IDPT system. The most common reasons cited for nonadherence in state-of-art studies were that (1) people believe they have made sufficient progress, (2) users reported there was too much content without much flexibility, (3) users reported that treatments were too complicated, (4) treatments did not match users' expectations, (5) there was a lack of therapist contact, and (6) there was a lack of personalization. A meta-analysis by Vandereycken and Devidt [20] concluded that the target groups chose not to adhere to the eating disorder treatment because they believed they had achieved sufficient

progress. However, a lack of progress is not related to nonadherence, according to several other studies [7,17]. According to a survey done by Johansson et al [21], participants chose not to adhere to the treatment when they were unable to perceive a compatible correlation between the length of weekly text modules and the conditions in their personal life. Moreover, the participants found the content to be a tiresome burden because of the length of the text modules and time consumed to go through them. Furthermore, the fixed format of the content sent to the participants each week was perceived as inflexible for some participants. Content complexity was perceived as challenging to comprehend and to process by individuals participating in interventions [8,21], especially when these individuals considered themselves to have attention problems or limited reading and writing skills. Participants' knowledge and expectations about the treatment process have shown to influence user trust and hence adherence [22]. Johansson et al [21] outlined in their study that participants mentioned they were grateful for being offered the treatment. However, not all of them appeared to be fully aware of the treatment and its significance. A similar conclusion was drawn by Alaoui et al [23], who identified higher treatment credibility to be the most influential prognostic factor for user adherence. Feedback has been thought to increase user adherence for 65% of intervention participants [19]. Similarly, a study by Johansson et al [21] revealed that the lack of therapist support during interventions was perceived by patients as a sign that therapists did not care about their health care issues. Furthermore, some participants reported that they never prioritized their personal development because they were aware that face-to-face meeting was not required. A recent study on mental health indicated that compliance failure can result from a lack of personalization [24]. A study by Doherty [25] claimed to have improved user adherence with the IDPT system by focusing on user personalization.

Most of the research examining the causes of low user adherence to IDPT has discovered that the reasons associated with patients were about personal and interpersonal competencies, and lack of resources rather than the diagnosis or health problem severity [7]. Moreover, it was about the patient's intrinsic motivation to change, their self-relatedness, and their receptivity to change. Levey and Clarkin [26] characterized this reason as the patient variable. Considering this as the reason for premature termination of interventions indicates a need to investigate the reasons and circumstances for nonadherence further. Specifically, this indicates a gap in the literature concerning the in-depth exploration of the subjective reasons for nonadherence in online psychological interventions. In general, the factors affecting premature termination of participants from IDPT, as outlined by Johansson et al [21], can be characterized by the interaction between the participant's perception of the treatment (content complexity, therapist feedback, and information about significance) and the participant's situation (awareness about the treatment, availability, daily routines, treatment expectations, and perceived language skills). Analogously, a report by the World Health Organization [27] distinguished five interacting dimensions affecting adherence to medication, therapy, and health care in general: socioeconomic factors, therapy-related factors,

patient-related factors, condition-related factors, and health system/health care team–related factors. The same report claimed that relatively limited research has been done on the effects of health system/health care team–related factors on adherence.

In this paper, we propose that in addition to these two factors (perception of treatment and personal situations), a third factor is contributing to user adherence: the adaptiveness of the IDPT system. There are two perspectives here: adaptiveness and information architecture (IA) [28]. First, IA is associated with how people cognitively process information and enhances the ability of the participants to find information. Second, adaptiveness refers to an ability in the system to change in response to environmental changes. The former perspective makes the information presented in IDPT comprehensible and discoverable, while the latter makes the IDPT more personalized. In this paper, we argue that both adaptiveness and IA are essential elements that contribute to user adherence in IDPT. Hence, we aim to investigate the following research questions in this literature review: (1) what are the most prevalent choices of IA in existing IDPT systems, and what is the primary rationale behind choosing an IA?, (2) what are the primary adaptive elements in IDPT systems, and how do these elements contribute to enhancing user adherence and intervention outcomes?, (3) what are the primary adaptive strategies used in IDPT systems, and how do these adaptive strategies consume adaptive elements to generate personalized experience for mental health patients?, and (4) how can we generalize the results to create a conceptual framework that can be used in the creation of an adaptive IDPT system for mental health interventions?

To the best of our knowledge, limited research has examined the experience of nonadherence in the IDPT system based on IA and adaptiveness as affecting factors. In this study, we focus on reviewing the adaptive elements and IA in the current IDPT systems used for the treatment of mental illness. Our review shows that several different terms are being used to describe similar IDPT systems. Interventions involving the internet as the communication mechanism are referred to as web-based treatments, web-based interventions, online treatment, computerized psychotherapy, e-therapy, eHealth, internet-based cognitive behavioral therapy, digital interventions, web app–based psychotherapy treatments, therapeutic web-based interventions, eHealth interventions [29], and others. Analogously, other variations include creation of technical platforms such as Interapy [30], Deprexis [31], ULTEMAT [32], digital behavior change interventions [33], and smartphone-based apps with specific brand names [34]. The absence of any taxonomic preferences and professional ontology makes the field of IDPT inconsistent and ambiguous. The use of a multitude of terms and labels to describe similar health interventions makes it difficult to search the results of studies. To be consistent, we chose to use the term IDPT, as suggested by Andersson et al [35].

Methods

We conducted the review according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [36]. Here, we present the methodology we used to search, analyze, and extract pertinent information from relevant studies.

Search Strategy

We searched the databases recommended by Cochrane [37], including Medline (PubMed), ACM Digital library, PsycINFO, EMBASE, CINAHL, and Cochrane, to identify studies. In addition, we hand-searched the reference list of the selected publications to retrieve additional relevant publications. The search string included Adaptive, OR Flexible, OR Tailored, AND Internet, AND Interventions, AND Mental Health (see Multimedia Appendix 1 for detailed search string). Each term included medical subject headings, and the search was done on full-text papers. The search was limited to all papers published in English from January 2000 to January 2020. The database searches and subsequent review were performed by the same two authors (SKM and JDW) independently in a double-blind process.

Eligibility Criteria

We included studies in which the articles met the following inclusion criteria: (1) discussed an intervention delivered through the internet (web- or mobile-based), (2) attempted to provide adaptive (dynamic, tailored, flexible) interventions by using adaptive strategies, (3) targeted a mental health disorder defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [38], and (4) was published between January 2000 and January 2020. No data restrictions were imposed. The following exclusion criteria were used: (1) not written in the English language, (2) not a full-text paper or published in the form of a short paper, extended abstract, abstract or poster, (3) designed as nonempirical findings such as opinion papers, reviews, editorials, letters, agendas, or study protocols, (4) dealt with adaptive technology in any domain other than mental health, or (5) was not about adaptive technology.

Review Procedure

The selection of studies took place in three phases based on the review of the title, keywords, abstract, and full text. Title and abstract screening were carried out blinded for author, journal, and date of publication. Any doubtful papers were included in the next phase, and disagreement was resolved through discussion. After identifying 3341 relevant papers in the initial database search, 372 duplicate papers were removed, and 2969 unique papers remained. In the screening step, the resulting list of 2969 papers was reviewed independently by the same two authors according to inclusion and exclusion criteria. By reviewing the title, abstract, and keywords, 105 eligible papers were retrieved. Two main reasons for the substantial exclusions were (1) the search engine returned the results containing any of the search terms, although they were logically connected, and (2) most of the papers were related to mental health without any reference to IDPT. Full texts were evaluated to determine

the eligibility of the remaining papers. The full texts of the 105 eligible papers were assessed independently by the same authors. Any discrepancies between the authors regarding the selection of the papers were resolved through discussion. In total, 74 papers were excluded in this round, and the selection process led to the inclusion of 31 papers, as illustrated in Figure 1. The most common reason for exclusion in this phase was that the publication did not discuss an intervention delivered via the internet. Other publications were excluded because they focused on other types of health care interventions without clear information about IA, user adherence, or adaptive strategies.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for this systematic review.
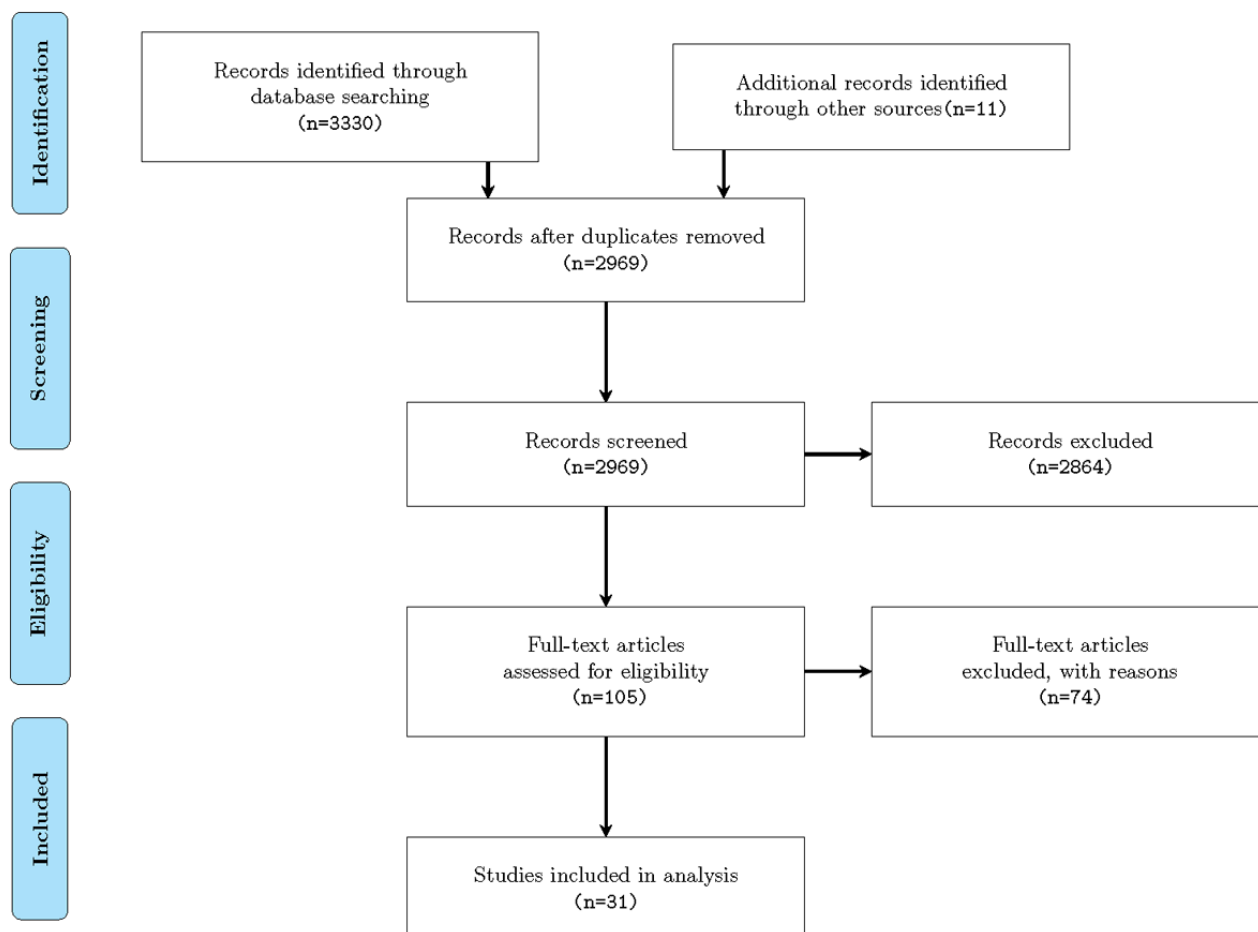
Data Extraction and Synthesis

Data from the included studies were extracted, verified, and tabulated for review by the authors. From the selected studies, we chose to obtain the main adaptive elements, adaptive strategies used, adaptive dimension, and actor involved in adaptation. Multimedia Appendix 2 provides a detailed summary of the analysis of the 31 articles in the review. All of the articles in the analysis are listed in the references [11-14,24,32,39-63]. We evaluated all of the relevant studies based on the adaptive IDPT model previously described [28]. As mentioned in the study [28], we extracted the core components of the adaptive reference model, including adaptive elements, adaptive dimensions, IA, and adaptive strategies. The rest of our results are based on these core components.

Data and Software Availability

For purposes of transparency and reproducibility of our study, we have published the resulting data, code, and procedures on GitHub [64]. The GitHub repository includes raw articles extracted from database searches, keyword formulation documents, preprocessed article lists, and a literate programming script used for data preprocessing, analysis, and visualization.

Results

Mental Health Illnesses Addressed

A significant number of the included studies addressed depression (n=11) and anxiety disorder (n=7), followed by general mental health issues (n=8), such as well-being, mindfulness, and goal achievement. Furthermore, some studies reported the use of adaptiveness in other areas such as insomnia (n=2), social psychology (n=1), attention deficit hyperactivity disorder (n=2), posttraumatic stress disorder (n=2), suicidality (n=2), and substance misuse (n=1). The full list of types of mental health problems addressed in the relevant studies is presented in Table 1.

Table 1. Types of mental illness for which an adaptive system was built.

Mental illnesses	Study references
Depression	Tsiakas et al, 2015 [40]
	Levin et al, 2018 [41]
	Burns et al, 2011 [50]
	Rebar et al, 2016 [51]
	Malins et al, 2020 [54]
	Van Gemert-Pijnen et al, 2014 [56]
	Lillevoll et al, 2014 [24]
	Achtyes et al, 2015 [57]
	Wallert et al, 2018 [58]
	Kop et al, 2014 [61]
	D'Alfonso et al, 2017 [63]
Anxiety disorder	Tsiakas et al, 2015 [40]
	Levin et al, 2018 [41]
	Walter et al, 2007 [45]
	Batterham et al, 2017 [48]
	Malins et al, 2020 [54]
	Achtyes et al, 2015 [57]
	Wallert et al, 2018 [58]
Insomnia	Forsell et al, 2019 [59]
	Erten-Uyumaz et al, 2019 [60]
Substance use	Batterham et al, 2017 [48]
Suicidality	Delgado-Gomez et al, 2016 [42]
	Batterham et al, 2017 [48]
Social psychology	Rachuri et al, 2010 [62]
Bipolar disorder	Dodd et al, 2017 [14]
Stress	Konrad et al, 2015 [47]
Posttraumatic stress disorder	Tielman et al, 2019 [43]
	Eisen et al, 2016 [46]
Smoking cessation	Lagoa et al, 2014 [44]
Attention deficit hyperactivity disorder	Nahum-Shani et al, 2012 [12]
General mental health	Iorfino et al, 2019 [39]
	Bannink et al, 2012 [11]
	Berrouiguet et al, 2018 [49]
	Ketelaar et al, 2014 [52]
	Coyle et al, 2010 [53]
	Kitagawa et al, 2020 [13]
	van Os et al, 2017 [55]
	van de Ven et al, 2017 [32]

Intervention Platform

Based on our findings, the communication media used to administer internet-facilitated interventions to patients can be classified into three categories: web apps, mobile apps, and computer games. A significant number of the included studies were based on web apps [11,14,24,39,43,49,50,52,54,56,63], followed by mobile apps [13,32,41,49,50,55,60,61] and a

game-based intervention [53]. There was only one paper that applied both web and mobile technologies for internet-based intervention. However, a lot of studies did not report the mode of delivery.

IA

IA is concerned with the art and science of organizing and labelling components of web apps, intranets, software, and online communities to enhance their usability and accessibility. IA plays a vital role in web app development, and a good architecture can improve the ability of employees and customers to find information and decrease the app's maintenance cost [28]. Finding the type of IA used in IDPT systems and their relevancy to treatment outcomes is one of the research questions of this study. However, a significant number of studies (20/31, 65%) did not report the type of IA they used in their IDPT system. Based on the IA of the intervention reported in the 31 reviewed articles, 4 (13%) studies reported the use of tunnel-based IA, and 3 (10%) studies reported the use of matrix IA and hierarchical IA. The full list of types of IA used in the relevant studies is presented in Table 2.

Table 2. Types of information architecture used in the reviewed studies.

Information architectures	Study references
Tunnel-based IA	Iorfino et al, 2019 [39]
	Konrad et al, 2015 [47]
	Batterham et al, 2017 [48]
	Kitagawa et al, 2020 [13]
Hybrid IA	D'Alfonso et al, 2017 [63]
Matrix IA	Levin et al, 2018 [41]
	Lagoa et al, 2014 [44]
	Van Gemert-Pijnen et al, 2014 [56]
Hierarchical IA	Tielman et al, 2019 [43]
	Bannink et al, 2012 [11]
	Berrouiguet et al, 2018 [49]
Not clear/not reported	Coyle et al, 2010 [53]
	Tsiakas et al, 2015 [40]
	Delgado-Gomez et al, 2016 [42]
	Walter et al, 2007 [45]
	Bannink et al, 2012 [11]
	Burns et al, 2011 [50]
	Rebar et al, 2016 [51]
	Ketelaar et al, 2014 [52]
	Malins et al, 2020 [54]
	Kitagawa et al, 2020 [13]
	van Os et al, 2017 [55]
	Lillevoll et al, 2014 [24]
	Dodd et al, 2017 [14]
	Achtyes et al, 2015 [57]
	Wallert et al, 2018 [58]
	Forsell et al, 2019 [59]
	Erten-Uyumaz et al, 2019 [60]
	Kop et al, 2014 [61]
	van de Ven et al, 2017 [32]
	Rachuri et al, 2010 [62]

The analysis of Table 2 answers our first research question. Most of these IAs fall into four categories: (1) tunnel-based design, (2) matrix design, (3) hierarchical design, and (4) hybrid design. A previous study [65] showed that 90% of the available

IDPT systems used a tunnel-based design, where users navigate sequentially to search for information. A tunnel-based design is analogous to watching TV series, reading textbooks, attending academic classes, or attending multiple clinical sessions. An argument for tunnel-based design is that the experience is less likely to overwhelm users with information and options [66]. The tunnel-based design is probably also the default IA design alternative in many projects and is the easiest to implement.

Adaptive Elements

Adaptive elements are the main components that are personalized for the user. As reported in a previous study [28], the main adaptive elements can be intervention content, design, assessment tests, IA, content presentation, content complexity, content recommendation, user interface (such as navigation system, search engines), feedback, notifications/reminders/alerts, behavioral activities, exercises, and reporting/dashboards. We report the full list of adaptive elements found in the relevant studies in Table 3.

Numerous studies (9/31, 29%) reported adapting the content of the intervention. However, most of these studies did not explicitly report the type of content, level of complexity, or modality (audio, video, presentation, pictures, assignments, activities, and assessments). Knowledge of the modalities of the content and their associated complexity provides insight into how interventions could be adapted and personalized for patients.

Another notable observation is that several studies (11/31, 35%) used feedbacks as adaptive elements. Numerous studies described the process of adaptive feedback in different forms, including sending personalized motivational messages [43], tailored messages by therapists [11,13,54], and providing general support [14]. In contrast, a few studies aimed to adapt reminders and alerts by sending an email or SMS text message or making a phone call [24,48,50]. Only 2 studies targeted the adaptation of exercises [41,47] and 1 study targeted the adaptation of behavioral activities [60]. We identified a total of 7 papers (7/31, 23%) that adapted assessment tests or psychometric assessment tests [39,42,45,46,55-57].

Table 3. Types of adaptive elements identified from the relevant studies.

Main adaptive elements	Study references
Intervention content	Iorfino et al, 2019 [39]
	Lagoa et al, 2014 [44]
	Batterham et al, 2017 [48]
	Rebar et al, 2016 [51]
	Coyle et al, 2010 [53]
	Nahum-Shani et al, 2012 [12]
	Van Gemert-Pijnen et al, 2014 [56]
	D'Alfonso et al, 2017 [63]
	Kop et al, 2014 [61]
Content presentation	Iorfino et al, 2019 [39]
Feedback message, support	Iorfino et al, 2019 [39]
	Tielman et al, 2019 [43]
	Bannink et al, 2012 [11]
	Batterham et al, 2017 [48]
	Burns et al, 2011 [50]
	Ketelaar et al, 2014 [52]
	Malins et al, 2020 [54]
	Kitagawa et al, 2019 [13]
	Van Gemert-Pijnen et al, 2014 [56]
	Dodd et al, 2017 [14]
	van de Ven et al, 2017 [32]
Assessment tests	Iorfino et al, 2019 [39]
	van Os et al, 2017 [55]
	Van Gemert-Pijnen et al, 2014 [56]
	Achtyes et al, 2015 [57]
	Delgado-Gomez et al, 2016 [42]
	Walter et al, 2007 [45]
	Eisen et al, 2016 [46]
Behavioral activities (sleep pattern)	Erten-Uyumaz et al, 2019 [60]
Reminder messages (SMS text messages, emails, phone calls)	Burns et al, 2011 [50]
	Lillevoll et al, 2014 [24]
	Batterham et al, 2017 [48]
Exercises	Levin et al, 2018 [41]
	Konrad et al, 2015 [47]
Reports	Iorfino et al, 2019 [39]
	Burns et al, 2011 [50]
Not clear	Tsiakas et al, 2015 [40]

Table 3 presents the list of the primary adaptive elements in the adaptive IDPT systems. The list includes (1) intervention content, (2) content presentation, (3) feedback messages, (4) assessment tests, (5) behavioral activities (sleep pattern), (6) reminder messages (email, SMS text messages, phone calls), (7) exercises, and (8) reporting (dashboard for the patients and the therapists). The central concept of adaptiveness is to create different levels of these adaptive elements and provide these elements based on a personalized profile. For example, if a person watches videos more than they listen to audio, read text, or view slides, then based on the principle of adaptiveness, it

makes sense to present upcoming interventions in a video format.

Dimensions of Adaptation

The way an adaptive system changes its behaviors depends on a multitude of factors: (1) users' data and preferences, (2) goals of the intervention, (3) measures, (4) adaptation actors, and (5) adaptation strategies. We refer to these aspects as the dimensions of the adaptive IDPT system [28]. Table 4 presents a list of adaptive dimensions extracted from the included studies.

Table 4. Dimensions considered for adaptation in the relevant studies.

Adaptation dimensions	Study references
User data and preferences (user context, needs, and location)	Iorfino et al, 2019 [39]
	Delgado-Gomez et al, 2016 [42]
	Tielman et al, 2019 [43]
	Lagoa et al, 2014 [44]
	Walter et al, 2007 [45]
	Eisen et al, 2016 [46]
	Van Gemert-Pijnen et al, 2014 [56]
	Dodd et al, 2017 [14]
	Forsell et al, 2019 [59]
	Erten-Uyumaz et al, 2019 [60]
	Kop et al, 2014 [61]
	van de Ven et al, 2017 [32]
	Rachuri et al, 2010 [62]
	D'Alfonso et al, 2017 [63]
Measures	
Psychometric tests/screening	Tsiakas et al, 2015 [40]
	Levin et al, 2018 [41]
	Bannink et al, 2012 [11]
	Batterham et al, 2017 [48]
	Berrouiguet et al, 2018 [49]
	Burns et al, 2011 [50]
	Rebar et al, 2016 [51]
	Ketelaar et al, 2014 [52]
	Coyle et al, 2010 [53]
	Malins et al, 2020 [54]
	Nahum-Shani et al, 2012 [12]
	Kitagawa et al, 2019 [13]
	Van Gemert-Pijnen et al, 2014 [56]
	Achtyes et al, 2015 [57]
	Wallert et al, 2018 [58]
User behavior analysis based on interaction data	Berrouiguet et al, 2018 [49]
	Burns et al, 2011 [50]
Goals of intervention	Konrad et al, 2015 [47]
Not clear	van Os et al, 2017 [55]
	Lillevoll et al, 2014 [24]

The relevant studies were mainly grouped into two clusters based on the choice of adaptive dimensions: user preferences (13/31, 42%) or outcome measures (14/31, 45%). Only 1 study used a goal-based adaptive dimension. Among the studies using

user preferences, some studies [32,42-46,62] used user context, while some studies used user location [32,62] to adapt interventions. The studies based on outcome measures used either psychometric tests or user behavior analysis based on interaction data to measure the performance outcome.

Adaptive Strategies

The adaptive strategy indicates the techniques used to tailor the intervention. In a recent study [28], four significant clusters of adaptive approaches were identified, namely rule-based adaptation, predictive algorithm-based (such as machine learning) adaptation, goal-driven adaptation, and adaptation through a feedback loop. Similar to the study [28], we identified the following adaptive strategies, presented in Table 5, in the reviewed studies.

Table 5. Types of adaptive strategies found in the relevant studies.

Types of adaptive strategies	Study references
Rule-based strategies	Iorfino et al, 2019 [39]
	Tsiakas et al, 2015 [40]
	Levin et al, 2018 [41]
	Delgado-Gomez et al, 2016 [42]
	Tielman et al, 2019 [43]
	Walter et al, 2007 [45]
	Eisen et al, 2016 [46]
	Bannink et al, 2012 [11]
	Konrad et al, 2015 [47]
	Batterham et al, 2017 [48]
	Rebar et al, 2016 [51]
	Ketelaar et al, 2014 [52]
	Coyle et al, 2010 [53]
	Malins et al, 2020 [54]
	Nahum-Shani et al, 2012 [12]
	Kitagawa et al, 2019 [13]
	van Os et al, 2017 [55]
	Van Gemert-Pijnen et al, 2014 [56]
	Lillevoll et al, 2014 [24]
	van de Ven et al, 2017 [32]
Predictive algorithm- or machine learning–based strategies	Tsiakas et al, 2015 [40]
	Lagoa et al, 2014 [44]
	Berrouiguet et al, 2018 [49]
	Burns et al, 2011 [50]
	Nahum-Shani et al, 2012 [12]
	Wallert et al, 2018 [58]
	Rachuri et al, 2010 [62]
	Erten-Uyumaz et al, 2019 [60]
	Kop et al, 2014 [61]
Recommendation-based strategies	D'Alfonso et al, 2017 [63]
General or unclear strategy	Dodd et al, 2017 [14]
	Achtyes et al, 2015 [57]

The list of strategies includes rule-based strategies, predictive algorithm-based strategies, and recommendation-based strategies. As expected, a significant number of studies (20/31, 65%) used some form of rule-based adaptation mechanism. For example, some studies [42,45,46] used item-response theory [8] to tailor psychometric tests. The primary motivation toward

adapting psychometric tests is to extract essential information from patients without asking them too many questions. The study by Van Gemert-Pijnen et al [56] applied user behavior analysis by inspecting web-log data and combined it with a rule-based engine to adapt the intervention. The same study concluded that pattern recognition can be a useful tool to tailor interventions based on usage patterns from earlier lessons [56]. With the hype of data science, several studies [12,40,44,49,50,58,60-62] attempted to use some form of predictive algorithm to adapt the intervention. While some studies did not report the outcome of the overall study [40,49], most of them reported that the use of predictive algorithms had a positive effect on the adaptation of interventions [12,44,50,58,60-62]. However, these studies concluded that further research is required to study the effectiveness of performance outcomes.

Efficacy of Treatment Outcomes

In general, this systematic review shows that tailoring interventions according to patients' needs and preferences has a positive effect on user adherence and hence treatment outcomes. Several studies reported that the personalization of interventions [13,32,39,41,43,47,50-52,54,56,59] and assessments [42,45,46] increased user adherence. Similarly, a study by Coyle et al [53] reported that a personalized system provided a higher degree of user satisfaction. However, some studies reported a lack of noticeable improvement in treatment outcomes. For instance, a study by Batterham et al [48] reported that there was no significant difference between tailored and static versions in the effectiveness of treatment or adherence. However, the same study reported that participants in the tailored conditions were more satisfied than those in the control conditions [48]. Lillevoll et al [24] made a similar conclusion, reporting that tailoring of feedback and dispatching weekly email reminders did not improve the intervention outcomes or user adherence. Given the scenario with two different clusters of results, further research and clinical trials are required to comprehend how user adherence and personalization of interventions are correlated.

Discussion

Key Findings

A total of 3341 studies were initially identified based on the inclusion criteria. Following a review of the title, abstract, and full text, 31 studies that fulfilled the inclusion criteria remained, most of which attempted to tailor interventions for mental illnesses. Approximately 68% (21/31) of the studies had a first author with a health care background. The most common adaptive elements were feedback messages to patients from therapists (11/31, 35%) and intervention content (9/31, 29%). However, how these elements contribute to the efficacy of IDPT in mental illness was not reported. The most common IA used was tunnel-based IA (4/31, 13%), while many studies (20/31, 65%) did not report the IA used. The rule-based technique was the most common adaptive strategy used in these studies (20/31, 65%). All the studies were broadly grouped into two adaptive dimensions based on user preferences or using performance measures such as psychometric tests.

Intervention Platform

Our findings show that web apps, mobile apps, and computer games are the primary platforms used to facilitate interventions. Apart from these, other communication media include robotics, virtual reality (VR) [67], augmented reality (AR), conversational agents, or chatbots [68]. A key finding in the literature study was that most of the IDPTs were made available on mobile apps [69] and web-based apps. This is as expected, as these are the most prevalent platforms used for personal computing. Smartphones contain a plethora of sensors and other data sources that inform aspects of users' well-being, context, activities, behaviors, and intentions. However, only a few attempts have been made to provide IDPT using conversational agents [68], or VR or AR apps. As depicted in Figure 2, higher placement in the chain indicates higher computational complexity but lower prevalence. Conversely, lower placement in the chain indicates lower computational complexity but higher prevalence. The most obvious explanation for the selection of web and mobile apps for intervention platforms is their prevalence. It makes the most sense to develop for platforms that are being used the most.

Figure 2. Mode of delivery for internet-delivered psychological treatments (IDPTs). AR: augmented reality; mHealth: mobile health; VR: virtual reality.
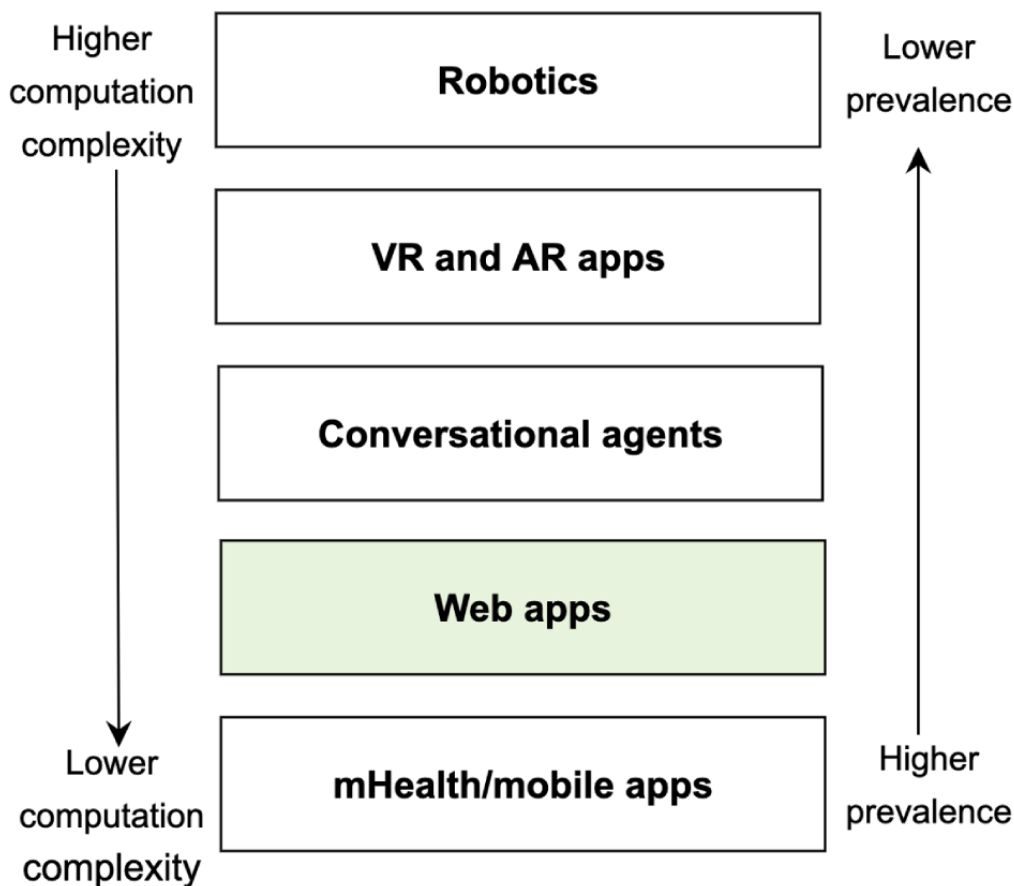
IA

Although a significant number of the studies failed to report which IA was used in their IDPT system, IA is still present in all software systems. Understanding the IA of a system helps a user to store, find, and interpret information readily, as IA is the design principle that is applied to making information discoverable and understandable. Finding the underlying IA of the IDPT system can help in making systems accessible and discoverable for end users, and knowledge about information design, structure, organization, and labelling can facilitate the development and evaluation phase. As explained in a previous study [70], IA consists of three major components: user, content, and context. Hence, IA helps to discover content for users based on their context. Assisting users with the correct piece of information they are looking for can increase user adherence by reducing bounce rate [71] and hence improve treatment outcomes. While our findings show most of the current adaptive IDPT systems are tunnel-based and don't promote personalization, we promote the use of hierarchical IA or hybrid IA in order to present useful information according to patients' needs and argue that the latter designs offer richer opportunities for adaptive IDPT systems.

Adaptive Elements

The primary context for building adaptive IDPT systems is to assist patients suffering from mental health disorders to learn about and recover from their illnesses. An IDPT system provides this information by using different media and elements such as text, video, audio, pictures, presentations, feedback, reminders, reports, and others. These elements have their format, structure, metadata, volume, and dynamism (such as frequency of updates). It is essential to understand these elements (contents described by Pakkala et al [71]) and the associated attributes in order to deliver the correct content to the right user in their current context. For example, understanding the complexity of the text describing sleep disorders can help to personalize the content based on the educational background of the patients. A general assumption is that an educated person can comprehend more complex language and medical terms than a nonspecialist. For a nonspecialist, it is more valuable to present the same content in terms of animated videos or pictures.

Adaptive Dimensions

The adaptive dimension provides the context for an adaptive IDPT system to tailor its behavior. A common way to tailor the behavior of the IDPT system is based on input regarding user preferences, measures (psychometric tests, user behavior analysis, and others), and goals, as shown in Table 4. In the context of user preferences, a study by Yardley et al [72] outlined that users can feel overwhelmed by the quantity and complexity of information presented to them. Hence, it is advisable to meet the needs of different users based on their needs and preferences. Personalization is critical mainly because users have different cognitive skills, educational backgrounds, content format preferences, comprehension capabilities, and

other qualities. A similar conclusion was made by Sundar and Marathe [73]. Their study identified two types of users: those influenced more by the affordance of agency (power users) and those influenced by the relevancy of the resulting content. The results of the study revealed that nonpower users preferred personalized content, whereas power users rated content quality higher when the website had a customizable feature. Similar to user preferences, other dimensions of adaption are performance measures such as the use of psychometric tests (eg, Patient Health Questionnaire-9 [74] for depression), user behavior analysis, and others.

Adaptive Strategies

Adaptive strategies provide a mechanism to present the right content to the right people based on their needs and preferences. While our review findings reveal that the rule-based adaptive strategy is the most widely adopted practice, other strategies, especially machine learning, are becoming highly prevalent. Given the premise that we can capture every digital footprint of a user, resulting in a complex and comprehensive data set, there is a possibility of using sophisticated machine learning or deep learning algorithms on the one hand, but it also raises an essential question about privacy on the other. In general, to build an adaptive IDPT system, it is crucial to understand which adaptive strategies can be used. Based on the selected strategies, one needs to collect and store the data. No matter which adaptive strategy is chosen, the adaption in an IDPT system is an iterative cycle where data is collected and preprocessed; preprocessed data are then analyzed and, based on the results of the analysis, an action is taken to tailor the intervention. However, how the data are analyzed and the result is extracted affect the way an IDPT system is developed, the choice of IA, the method of data storage, and other parameters.

Challenges in Contemporary IDPT Systems

Although the integration of health care systems has emerged as a policy for several health care agencies, there is a large gap between current policy, program implementation efforts, and evidence for health care integration. The results of our review led us to list the following challenges in current IDPT practice.

Lack of Standard Taxonomy in IDPT

There is a lack of a standardized definition of the health care system and proper taxonomy to allow the grouping of similar interventions. As mentioned in the introduction, the use of nonstandard terms to refer to the same system causes inconsistencies and makes it hard to draw conclusions. Based on this challenge, several researchers [75-77] have made an effort to formalize the health care system to support interoperability [76], such as the Fast Healthcare Interoperability Resources (FHIR) created by the Health Level Seven International (HL7) health care standards organization [78] and others. If there is a standard followed by researchers for building adaptive IDPT systems, it will become easier to learn from their findings and extend the current understanding to improve treatment outcomes.

Scientific Foundation

The outcome of trials of IDPT systems has demonstrated comparable results as face-to-face therapies. However, despite considerable attention to IDPTs, user adherence is low, and there is remarkably less literature on the underlying science of the field of IDPT system design and development [79]. Many pieces of literature claim IDPT to be based on psychoeducation that helps in the modification of behavior change and symptom improvement. While this assumption may be correct, the underlying science behind how psychoeducation about particular symptoms enhances behavior modifications and symptom improvement is less evident in the literature. Lack of a scientific foundation behind IDPT systems may be the reason behind users' lack of trust toward interventions [21-23] and hence lack of adherence. In the same way, an IDPT system is an application software that follows an IA (see Table 2) and design patterns [80]. The application software such as IDPT are well formalized and studied in the research community. Lack of such reporting makes it challenging to conclude how adaptive elements or the IA influence the outcome of an intervention.

Ethical and Safety Issues Associated with Predictive Adaptive Strategies

Technology has matured to the point where several researchers envision the creation of automated, adaptive IDPT systems that work without much human involvement. However, there are controversies between what is possible and what is acceptable in adaptive systems. Hence, it requires careful consideration of both ethical and legal issues; focusing solely on technological and operational perspectives can lead to low value or utility for patients. As a result, both information and communication technology (ICT) researchers and medical practitioners must consider the capabilities, limitations, and needs of patients when designing adaptive systems. The primary objective of the adaptive IDPT system is to tailor the intervention based on user needs or any other adaptive dimensions. The adaptive IDPT system can understand the user's needs by creating detailed user profiling. User profiling includes storage of the patient's previous diagnosis, sensitive personal information, as well as the current status. Moreover, to maximize the benefits of data-driven adaptiveness, the adaptive IDPT system needs to store interaction data, including the time of login, the frequency of login, and the interaction with the system at the granular level (clicks, keystrokes). For example, the study by Van Gemert-Pijnen [56] analyzed the log data in order to understand the use of the content. Storing such user interaction data requires proper user consent on the one hand and directly deals with the privacy of the patient on the other. Hence, it is one of the critical challenges in the development of an adaptive IDPT system. It requires further research into the problem of storing user interaction data securely. For example, many psychological interventions aim to characterize patients' symptoms based on their mobile phone usage. This type of study is possible because mobile phones come with built-in sensors and standard application programming interfaces to measure and collect patients' data, including mobility patterns, physical activities, crowd density, time spent indoors versus outdoors, and locations. Although these capabilities are possible with the advances in ubiquitous computing, they deal directly with privacy and ethical issues.

Implications and Future Directions

It is not easy to predict how technologies will develop over time and whether these technologies will continue adapting to clinical use. However, based on the results of this systematic review, we outline some implications and future directions in the field of IDPT system development and innovation.

Implications for ICT Researchers

With an increasing trend in user adherence toward internet-delivered treatments on the one hand and the prevalence of the internet of things (IoT), with growth in ambient intelligence technology, on the other, there is an expectation that the IDPT system will flourish over time. A plethora of health care interventions delivered via the internet have a similar format, as most of them are based on psychoeducation. All such interventions attempt to create adaptive elements (see Table 3) and attempt to tailor these elements based on adaptive dimensions (see Table 4) using adaptive strategies (see Table 5). Hence, it makes sense to create a conceptual framework that can be utilized in several health care domains. Moreover, the creation of domain-specific languages for incorporating adaptive health care interventions is also required. We also need better dashboard tools that help therapists and other medical practitioners to comprehend patients' status better and adapt their interventions based on their engagement with the interventions.

We analyzed the state-of-the-art studies concerned with adapting psychological interventions. The analysis yielded the answers to the most critical questions, including (1) what are the essential elements that therapists wish to tailor? (2) what are the main dimensions in which these elements can be tailored to meet patients' needs? and (3) what are the primary adaptive strategies used to trigger adaptation in those dimensions? Findings from the analysis helped to identify the essential variables that are associated with an adaptive system. As McGaghie et al [81] outlined, once ICT researchers and developers know the essential variables, they can utilize these findings to create a conceptual framework that sets the stage for the presentation of a particular problem, which in this case is the creation of an adaptive IDPT system. Further, ICT researchers and developers can validate the conceptual framework by building domain-specific language.

Implication for Health Care Researchers

While the current research evidence is fragmented about the benefit of an adaptive IDPT system on treatment outcomes, this review suggests that adaptive IDPT systems can benefit people with mental health issues in providing personalized psychoeducation. Such an education will help mental health patients to manage their illness. In addition, a high number of health care researchers have published about adaptive interventions, as shown in Figure 2. This indicates that both health care researchers and computer researchers believe that adaptive interventions are an essential phenomenon to accelerate user adherence. However, tailoring the feedbacks given by therapists or providing reminders are simple forms of adaptation (see Table 3). The intervention can be adapted for several dimensions, including user preferences, outcome measures, and

different adaptive strategies [28], with the amalgamation of ambient technology. Hence, cooperation between ICT and health care researchers is essential to develop an adaptive IDPT system. While more randomized controlled trials are required to validate the effectiveness of adaptive treatments, the results of this review show sufficient evidence to suggest that the adaptation of mental health interventions can enhance user adherence and treatment outcomes.

Implication for Computer Science Research

The development of an adaptive IDPT system that increases user adherence and treatment outcomes requires more extensive research to establish clinical appropriateness. Given the potential benefit of the IDPT system for cost-effective delivery to the far-reaching population, further research should be conducted on how to personalize adaptive strategies. Furthermore, reporting back to the research community is the part of any discipline of transparency that keeps studies honest and accountable. In addition, it fits into the broader responsibilities of scientists to communicate their work and foster public understanding. Such understanding can be used by other researchers to gain insight into new research directions.

Future Work

An immediate future task involves the creation of a conceptual framework for adaptive IDPT systems. In addition to this, we envision the development of domain-specific language that can model such an adaptive IDPT system. Furthermore, it is imperative from the review that there is a need for a comprehensive visual dashboard for therapists and patients where they can receive the intervention, monitor their symptoms, and manage their illness.

Limitations

Given that the health ICT literature is quite diverse and extensive, the current study focused exclusively on internet-delivered interventions for mental health morbidities. Notwithstanding this limitation, this paper highlights the significance of the continued study of this intervention method. Another limitation is that our literature exploration only encompassed articles in the English language; therefore, it is plausible that some research conducted in other parts of the world and published in other languages were missed. A third limitation pertains to IDPT apps developed by industry that were not accessible for review. Hence, we have less knowledge about the adaptive elements involved in their architecture.

Conclusions

Adaptive psychological interventions tailor the type of content or tasks to individuals based on their needs and preferences in order to improve saliency and intervention efficacy. This systematic review describes the investigation and analysis of existing studies about adaptive psychological intervention delivered through the internet. The study outlines the main elements used in the process of adaptation, the IA used in the adaptive systems, the main dimensions of adaptation, and the main adaptive strategies. Based on these findings, we envision the development of a conceptual framework that researchers and clinicians can utilize to build adaptive models of several health care interventions.

The findings of our review indicate the use of web-based and mobile apps to deliver mental health interventions, such as for depression (most studied), anxiety, and others. However, a number of these studies did not report the IA used in their system, and those that did report mostly used tunnel-based systems. Similarly, several studies used rule-based adaptive strategies to adapt intervention based on performance measures such as psychometric tests. Feedback messages, reminders, and support were the most used adaptive element. Further study is required to explore the role of IA, adaptive elements, adaptive dimensions, and adaptive strategies in building a successful IDPT system. Knowledge about these core elements of the adaptive IDPT system can serve to create a conceptual framework that can be used for several health care interventions.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms that were used to execute the search in different databases.
[PDF File (Adobe PDF File), 90 KB - jmir_v22i11e21066_app1.pdf]

Multimedia Appendix 2
The main table that helped to extract the adaptive elements, adaptive strategies, information architecture, and other information about the internet-delivered psychological treatment system.
[XLSX File (Microsoft Excel File), 68 KB - jmir_v22i11e21066_app2.xlsx]

References

1. Portnoy D, Scott-Sheldon LAJ, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]

2. Marks IM, Cavanagh K, Gega L. Computer-aided psychotherapy: revolution or bubble? Br J Psychiatry 2018 Jan 02;191(6):471-473 [FREE Full text] [doi: 10.1192/bjp.bp.107.041152]

3. Wallin EEK, Mattsson S, Olsson EMG. The Preference for Internet-Based Psychological Interventions by Individuals Without Past or Current Use of Mental Health Treatment Delivered Online: A Survey Study With Mixed-Methods Analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]

4. Bennett G, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292 [FREE Full text] [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

5. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

6. Dunn T, Casey LM, Sheffield J, Newcombe P, Chang AB. Dropout from computer-based interventions for children and adolescents with chronic health conditions. J Health Psychol 2012 Apr;17(3):429-442 [FREE Full text] [doi: 10.1177/1359105311415558] [Medline: 21890540]

7. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

8. Fraley R, Waller N, Brennan K. An item response theory analysis of self-report measures of adult attachment. Journal of Personality and Social Psychology 2000;78(2):350-365 [FREE Full text] [doi: 10.1037/0022-3514.78.2.350]

9. Andersson G. Using the Internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180 [FREE Full text] [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]

10. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar 20;37(3):319-328 [FREE Full text] [doi: 10.1017/S0033291706008944] [Medline: 17112400]

11. Bannink R, Joosten-van Zwanenburg E, van de Looij-Jansen P, van As E, Raat H. Evaluation of computer-tailored health education ('E-health4Uth') combined with personal counselling ('E-health4Uth + counselling') on adolescents' behaviours and mental health status: design of a three-armed cluster randomised controlled trial. BMC Public Health 2012 Dec 17;12:1083 [FREE Full text] [doi: 10.1186/1471-2458-12-1083] [Medline: 23244449]

12. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Q-learning: a data analysis method for constructing adaptive interventions. Psychol Methods 2012 Dec;17(4):478-494 [FREE Full text] [doi: 10.1037/a0029373] [Medline: 23025434]

13. Kitagawa T, Higuchi Y, Todo E, Ueda T, Ando S, Murakami T. Tailored feedback reduced prolonged sitting time and improved the health of housewives: a single-blind randomized controlled pilot study. Women & Health 2019 May 21;60(2):212-223 [FREE Full text] [doi: 10.1080/03630242.2019.1616043]

14. Dodd A, Mallinson S, Griffiths M, Morriss R, Jones SH, Lobban F. Users' experiences of an online intervention for bipolar disorder: important lessons for design and evaluation. Evid Based Ment Health 2017 Nov;20(4):133-139 [FREE Full text] [doi: 10.1136/eb-2017-102754] [Medline: 29061886]

15. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]

16. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]

17. Klein B, Austin D, Pier C, Kiropoulos L, Shandley K, Mitchell J, et al. Internet-based treatment for panic disorder: does frequency of therapist contact make a difference? Cogn Behav Ther 2009;38(2):100-113 [FREE Full text] [doi: 10.1080/16506070802561132] [Medline: 19306149]

18. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468 [FREE Full text] [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]

19. Bewick B, Trusler K, Mulhern B, Barkham M, Hill AJ. The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: a randomised control trial. Addict Behav 2008 Sep;33(9):1192-1198 [FREE Full text] [doi: 10.1016/j.addbeh.2008.05.002] [Medline: 18554819]

20. Vandereycken W, Devidt K. Dropping out from a specialized inpatient treatment for eating disorders: the perception of patients and staff. Eat Disord 2010;18(2):140-147 [FREE Full text] [doi: 10.1080/10640260903585557] [Medline: 20390617]

21. Johansson O, Michel T, Andersson G, Paxling B. Experiences of non-adherence to Internet-delivered cognitive behavior therapy: A qualitative study. Internet Interventions 2015 May;2(2):137-142 [FREE Full text] [doi: 10.1016/j.invent.2015.02.006]

22. Sledge W, Moras K, Hartley D, Levine M. Effect of time-limited psychotherapy on patient dropout rates. Am J Psychiatry 1990 Oct;147(10):1341-1347 [FREE Full text] [doi: 10.1176/ajp.147.10.1341] [Medline: 2400003]

23. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of Symptomatic Change and Adherence in Internet-Based Cognitive Behaviour Therapy for Social Anxiety Disorder in Routine Psychiatric Care. PLoS One 2015;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]

24. Lillevoll K, Vangberg HCB, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]

25. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: Association for Computing Machinery, New York, NY, United States; 2012 May Presented at: In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May, 2012; Austin Texas USA p. 1421-1430 URL: https://doi.org/10.1145/2207676.2208602 [doi: 10.1145/2207676.2208602]

26. Garfield SL, Bergin AE, Dryden W. Handbook of Psychotherapy and Behavior Change. The influence of client variables on psychotherapy. Bergin and Garfield?s 1987 Jan 01;1(4):264.2-26265. [doi: 10.1891/0889-8391.1.4.264a]

27. World Health Organization, WHO. Adherence to long-term therapies: evidence for action. In: In ADHERENCE TO LONG-TERM THERAPIES: Evidence for action (p. 211). USA: World Health Organization; 2003.

28. Mukhiya SK, Wake JD, Inal Y, Lamo Y. Adaptive Systems for Internet-Delivered Psychological Treatments. IEEE Access 2020;8:112220-112236. [doi: 10.1109/access.2020.3002793]

29. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: Outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interventions 2015 Mar;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002]

30. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PMG. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]

31. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009 May 11;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]

32. van de Ven P, O'Brien H, Henriques R, Klein M, Msetfi R, Nelson J, E-COMPARED Consortium. ULTEMAT: A mobile framework for smart ecological momentary assessments and interventions. Internet Interv 2017 Sep;9:74-81 [FREE Full text] [doi: 10.1016/j.invent.2017.07.001] [Medline: 30135840]

33. Pejovic V, Musolesi M. Anticipatory mobile computing for behaviour change interventions. In: UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. New York, NY: Association for Computing Machinery; 2014 Sep Presented at: Anticipatory mobile computing for behaviour change interventions. In UbiComp - Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 1025?1034). Association for Computing Machinery, Inc; 2014; Seattle Washington, USA p. 1025-1034 URL: https://doi.org/10.1145/2638728.2641284 [doi: 10.1145/2638728.2641284]

34. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

35. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]

36. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

37. Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board, Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976) 2009 Aug 15;34(18):1929-1941. [doi: 10.1097/BRS.0b013e3181b1c99f] [Medline: 19680101]

38. American Psychiatric Association, APA. American Psychiatric Association, 2013. In: Diagnostic and statistical manual of mental disorders (5th ed). USA: American Psychiatric Association; May 2013.

39. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A Digital Platform Designed for Youth Mental Health Services to Deliver Personalized and Measurement-Based Care. Front Psychiatry 2019;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]

40. Tsiakas K, Watts L, Lutterodt C, Giannakopoulos T, Papangelis A, Gatchel R, et al. A multimodal adaptive dialogue manager for depressive and anxiety disorder screening: a Wizard-of-Oz experiment. In: PETRA '15: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2015 Jul Presented at: PETRA '15: Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments; July, 2015; Corfu Greece p. 1-4 URL: https://doi.org/10.1145/2769493.2769572 [doi: 10.1145/2769493.2769572]

41. Levin M, Navarro C, Cruz RA, Haeger J. Comparing in-the-moment skill coaching effects from tailored versus non-tailored acceptance and commitment therapy mobile apps in a non-clinical sample. Cogn Behav Ther 2019 May;48(3):200-216 [FREE Full text] [doi: 10.1080/16506073.2018.1503706] [Medline: 30117784]

42. Delgado-Gomez D, Baca-Garcia E, Aguado D, Courtet P, Lopez-Castroman J. Computerized Adaptive Test vs. decision trees: Development of a support decision system to identify suicidal behavior. J Affect Disord 2016 Dec;206:204-209 [FREE Full text] [doi: 10.1016/j.jad.2016.07.032] [Medline: 27475891]

43. Tielman M, Neerincx MA, Brinkman WP. Design and Evaluation of Personalized Motivational Messages by a Virtual Agent that Assists in Post-Traumatic Stress Disorder Therapy. J Med Internet Res 2019 Mar 27;21(3):e9240 [FREE Full text] [doi: 10.2196/jmir.9240] [Medline: 30916660]

44. Lagoa C, Bekiroglu K, Lanza ST, Murphy SA. Designing adaptive intensive interventions using methods from engineering. J Consult Clin Psychol 2014 Oct;82(5):868-878 [FREE Full text] [doi: 10.1037/a0037736] [Medline: 25244394]

45. Walter O, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). Qual Life Res 2007;16 Suppl 1:143-155. [doi: 10.1007/s11136-007-9191-7] [Medline: 17342455]

46. Eisen S, Schultz MR, Ni P, Haley SM, Smith EG, Spiro A, et al. Development and Validation of a Computerized-Adaptive Test for PTSD (P-CAT). Psychiatr Serv 2016 Oct 01;67(10):1116-1123 [FREE Full text] [doi: 10.1176/appi.ps.201500382] [Medline: 27247175]

47. Konrad A, Bellotti V, Crenshaw N, Tucker S, Nelson L, Du H, et al. Finding the Adaptive Sweet Spot: Balancing Compliance and Achievement in Automated Stress Reduction. New York, NY: Association for Computing Machinery; 2015 Apr Presented at: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April, 2015; Seoul Republic of Korea p. 3829-3838 URL: https://doi.org/10.1145/2702123.2702512 [doi: 10.1145/2702123.2702512]

48. Batterham P, Calear AL, Farrer L, McCallum SM, Cheng VWS. : Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]

49. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: Transition to Participatory and Personalized Medicine in Mental Health. J Med Internet Res 2018 Jan 03;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]

XSL•FO

RenderX

50. Burns M, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]

51. Rebar A, Boles C, W. Burton N, Duncan M, Short C, Happell B, et al. Healthy mind, healthy body: A randomized trial testing the efficacy of a computer-tailored vs. interactive web-based intervention for increasing physical activity and reducing depressive symptoms. Mental Health and Physical Activity 2016 Oct;11:29-37 [FREE Full text] [doi: 10.1016/j.mhpa.2016.08.001]

52. Ketelaar S, Nieuwenhuijsen K, Bolier L, Smeets O, Sluiter JK. Improving work functioning and mental health of health care employees using an e-mental health approach to workers' health surveillance: pretest-posttest study. Saf Health Work 2014 Dec;5(4):216-221 [FREE Full text] [doi: 10.1016/j.shaw.2014.08.002] [Medline: 25516815]

53. Coyle D, Doherty G, Sharry J. PlayWrite: end-user adaptable games to support adolescent mental health. New York, NY: Association for Computing Machinery; 2010 Apr Presented at: CHI EA '10: CHI '10 Extended Abstracts on Human Factors in Computing Systems; April, 2010; Atlanta Georgia USA p. 3889-3894 URL: https://doi.org/10.1145/1753846.1754074 [doi: 10.1145/1753846.1754074]

54. Malins S, Biswas S, Patel S, Levene J, Moghaddam N, Morriss R. Preventing relapse with personalized smart-messaging after cognitive behavioural therapy: A proof-of-concept evaluation. Br J Clin Psychol 2020 Jun 20;59(2):241-259 [FREE Full text] [doi: 10.1111/bjc.12244] [Medline: 31960467]

55. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493 [FREE Full text] [doi: 10.1002/da.22647] [Medline: 28544391]

56. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]

57. Achtyes E, Halstead S, Smart L, Moore T, Frank E, Kupfer DJ, et al. Validation of Computerized Adaptive Testing in an Outpatient Nonacademic Setting: The VOCATIONS Trial. Psychiatr Serv 2015 Oct;66(10):1091-1096 [FREE Full text] [doi: 10.1176/appi.ps.201400390] [Medline: 26030317]

58. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting Adherence to Internet-Delivered Psychotherapy for Symptoms of Depression and Anxiety After Myocardial Infarction: Machine Learning Insights From the U-CARE Heart Randomized Controlled Trial. J Med Internet Res 2018 Oct 10;20(10):e10754 [FREE Full text] [doi: 10.2196/10754] [Medline: 30305255]

59. Forsell E, Jernelöv S, Blom K, Kraepelien M, Svanborg C, Andersson G, et al. Proof of Concept for an Adaptive Treatment Strategy to Prevent Failures in Internet-Delivered CBT: A Single-Blind Randomized Clinical Trial With Insomnia Patients. Am J Psychiatry 2019 Apr 01;176(4):315-323 [FREE Full text] [doi: 10.1176/appi.ajp.2018.18060699] [Medline: 30696270]

60. Erten-Uyumaz B, Ahuja M, Vacaretu T, Rama MD, Overeem S, Visser T, et al. Design and evaluation of a negotiation-based sleep scheduler app for insomnia treatment. New York, NY: Association for Computing Machinery; 2019 May Presented at: Pervasive Health'19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May, 2019; Trento Italy p. 225-233 URL: https://doi.org/10.1145/3329189.3329218 [doi: 10.1145/3329189.3329218]

61. Kop R, Hoogendoorn M, Klein CAM. A Personalized Support Agent for Depressed Patients: Forecasting Patient Behavior Using a Mood and Coping Model. : 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT); 2014 Presented at: IEEE WIC ACM International Conference on Web Intelligence (WI); 11-14 Aug. 2014; Warsaw, Poland URL: https://doi.org/10.1109/WI-IAT.2014.181 [doi: 10.1109/wi-iat.2014.181]

62. Rachuri K, Musolesi M, Mascolo C, Rentfrow P, Longworth C, Aucinas A. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. New York, NY: Association for Computing Machinery; 2010 Sep Presented at: UbiComp '10: Proceedings of the 12th ACM international conference on Ubiquitous computing; September 2010; Copenhagen Denmark p. 281-290 URL: https://doi.org/10.1145/1864349.1864393 [doi: 10.1145/1864349.1864393]

63. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health. Front Psychol 2017;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

64. Supplement resources in GitHub. GitHub. URL: https://github.com/sureshHARDIYA/phd-resources/tree/master/Data/Review%20Paper [accessed 2020-11-15]

65. Kelders S, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

66. Pugatch J, Grenen E, Surla S, Schwarz M, Cole-Lewis H. Information Architecture of Web-Based Interventions to Improve Health Outcomes: Systematic Review. J Med Internet Res 2018 Mar 21;20(3):e97 [FREE Full text] [doi: 10.2196/jmir.7867] [Medline: 29563076]

67. Kang N, Brinkman WP, Riemsdijk MBV, Neerincx MA. Internet-delivered multi-patient virtual reality exposure therapy system for the treatment of anxiety disorders. In: ECCE '11: Proceedings of the 29th Annual European Conference on

Cognitive Ergonomics. New York, NY: Association for Computing Machinery; 2011 Aug Presented at: In ECCE - European Conference on Cognitive Ergonomics the Annual Conference of the European Association of Cognitive Ergonomics (pp. 233?236); August, 2011; Rostock Germany p. 233-236 URL: https://doi.org/10.1145/2074712.2074760 [doi: 10.1145/2074712.2074760]

68. Fitzpatrick K, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

69. Johnston L, Titov N, Andrews G, Spence J, Dear BF. A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]

70. Holtzblatt K, Beyer H. Contextual design: using customer work models to drive systems design. In: CHI '98: CHI 98 Conference Summary on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 1998 Apr Presented at: CHI '98: CHI 98 Conference Summary on Human Factors in Computing Systems; April, 1998; Los Angeles California USA p. 149-150. [doi: 10.1145/286498.286629]

71. Pakkala H, Presser K, Christensen T. Using Google Analytics to measure visitor statistics: The case of food composition websites. International Journal of Information Management 2012 Dec;32(6):504-512 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2012.04.008]

72. Yardley L, Morrison LG, Andreou P, Joseph J, Little P. Understanding reactions to an internet-delivered health-care intervention: accommodating user preferences for information provision. BMC Med Inform Decis Mak 2010 Sep 17;10:52 [FREE Full text] [doi: 10.1186/1472-6947-10-52] [Medline: 20849599]

73. Sundar SS, Marathe SS. Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. Human Communication Research 2010 Jul 01;36(3):298-322 [FREE Full text] [doi: 10.1111/j.1468-2958.2010.01377.x]

74. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov 14;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]

75. Porterfield D, Rojas-Smith L, Lewis M. A taxonomy of integration interventions between health care and public health. In: RTI Press. Research Triangle Park (NC): RTI Press; Jul 2015.

76. Mukhiya S, Rabbi F, I Pun V, Rutle A, Lamo Y. A GraphQL approach to Healthcare Information Exchange with HL7 FHIR. Procedia Computer Science 2019;160:338-345 [FREE Full text] [doi: 10.1016/j.procs.2019.11.082]

77. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. 2013 Presented at: Annual IEEE Symposium on Computer-Based Medical Systems; 20-22 June 2013; Porto, Portugal p. 326-331 URL: https://doi.org/10.1109/CBMS.2013.6627810 [doi: 10.1109/cbms.2013.6627810]

78. Suresh KM, Fazle R, Ka IP, Yngve L. An Architectural Design for Self-Reporting E-Health Systems. In: IEEE/ACM International Workshop on Software Engineering for Healthcare (SEH). 2019 Presented at: An architectural design for self-reporting e-health systems. In Proceedings of the 1st International Workshop on Software Engineering for Healthcare (SEH ?19). IEEE Press, 1?8; 2019; Montreal, QC, Canada p. 1-8 URL: https://doi.org/10.1109/SEH.2019.00008 [doi: 10.1109/seh.2019.00008]

79. Ritterband L, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3 [FREE Full text] [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]

80. Gamma E, Helm R, Johnson R, Vlissides J. Design Patterns: Abstraction and Reuse of Object-Oriented Design. In: Broy M, Denert E, editors. Pioneers and Their Contributions to Software Engineering. Heidelberg: Springer, Berlin, Heidelberg; 2001.

81. McGaghie W, Bordage G, Shea J. Problem Statement, Conceptual Framework, and Research Question. Academic Medicine 2001;76(9):923-924 [FREE Full text] [doi: 10.1097/00001888-200109000-00021]

Abbreviations

AR: augmented reality
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven International
IA: information architecture
ICT: information communication technology
IDPT: internet-delivered psychological treatment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality

Viewpoint

The COVID-19 Pandemic: A Pandemic of Lockdown Loneliness and the Role of Digital Technology

Syed Ghulam Sarwar Shah[1,2], MBBS, MA, MSc, PhD; David Nogueras[3], MBA; Hugo Cornelis van Woerden[4,5,6], MBChB, MPH, FFPH, MRCGP, PhD; Vasiliki Kiparoglou[1,7], BSc, MSc, MBA, PhD

[1]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[2]Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

[3]EvZein Limited, Oxford, United Kingdom

[4]Public Health Agency Northern Ireland, Belfast, United Kingdom

[5]Division of Rural Health and Wellbeing, University of the Highlands and Islands, Inverness, United Kingdom

[6]Institute of Nursing and Health Research, Ulster University, Belfast, United Kingdom

[7]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Syed Ghulam Sarwar Shah, MBBS, MA, MSc, PhD
NIHR Oxford Biomedical Research Centre
Oxford University Hospitals NHS Foundation Trust
John Radcliffe Hospital
Oxford, OX3 9DU
United Kingdom
Phone: 44 10865 221262
Email: sarwar.shah@ouh.nhs.uk

Abstract

The focus of this perspective is on lockdown loneliness, which we define as loneliness resulting from social disconnection as a result of enforced social distancing and lockdowns during the COVID-19 pandemic. We also explore the role of digital technology in tackling lockdown loneliness amid the pandemic. In this regard, we highlight and discuss a number of the key relevant issues: a description of lockdown loneliness, the burden of lockdown loneliness during the COVID-19 pandemic, characteristics of people who are more likely to be affected by lockdown loneliness, factors that could increase the risk of loneliness, lockdown loneliness as an important public health issue, tackling loneliness during the pandemic, digital technology tools for social connection and networking during the pandemic, assessment of digital technology tools from the end users' perspectives, and access to and use of digital technology for tackling lockdown loneliness during the COVID-19 pandemic. We suggest that the most disadvantaged and vulnerable people who are more prone to lockdown loneliness are provided with access to digital technology so that they can connect socially with their loved ones and others; this could reduce loneliness resulting from social distancing and lockdowns during the COVID-19 crisis. Nonetheless, some key issues such as access to and knowledge of digital technology tools must be considered. In addition, the involvement of all key stakeholders (family and friends, social care providers, and clinicians and health allied professionals) should be ensured.

Background

The COVID-19 pandemic has swept across the globe, resulting in about 29.3 million confirmed cases and about 0.93 million deaths worldwide as of September 15, 2020 [1]. The pandemic has compelled governments and authorities in affected countries to enforce preventive measures including enforced lockdowns, social distancing, self-isolation, and quarantine to slow down the spread of COVID-19 [2]. These preventative measures have contributed to social isolation and loneliness among people with specific characteristics [3,4]. The current COVID-19 crisis has profoundly affected social connections, and digital technology is playing an important role by providing virtual opportunities not only for businesses and health care delivery but also for

social connection and networking. Numerous digital technology tools for social connection and business are available and being used by individuals according to their specific needs and requirements. However, certain groups of people are more disadvantaged because they do not have access to these tools and do not have the resources to get them. Hence, these underprivileged people are more likely to be socially disconnected and are at greater risk of loneliness, which is associated with serious adverse effects on social, physical, and mental health [4,5]. In this context, we identify and discuss a number of relevant and important issues as follows.

What is Loneliness?

Loneliness is a subjective feeling of perceived "mismatch between the quantity and quality of social relationships" [6,7]. Loneliness is also commonly reported as a perceived discrepancy between the actual and desired social relationships of an individual [8].

A recent study on loneliness during the COVID-19 pandemic measured and reported loneliness as "chronic loneliness" (feeling lonely often or always) and "lockdown loneliness" (feeling lonely during the past 7 days) [3].

We consider lockdown loneliness more germane to, and of much interest during, the COVID-19 pandemic. We therefore focus on lockdown loneliness, which we define as "loneliness resulting because of social disconnection due to enforced social distancing and lockdowns during the COVID-19 pandemic and similar other emergency situations."

Burden of Lockdown Loneliness During the COVID-19 Pandemic

Although a loneliness epidemic [9] was reported in many countries (including Australia, the United Kingdom, and the United States) prior to the COVID-19 pandemic [10], the burden of loneliness has increased during pandemic lockdowns [11]. An increase in lockdown loneliness during the COVID-19 pandemic is evident from the latest statistics on coronavirus and loneliness in Great Britain, released in June 2020, which show that lockdown loneliness affected about 7.4 million adults (equivalent to about 14% of residents) during the COVID-19 pandemic lockdowns, while chronic loneliness remained at similar levels compared to prelockdown (2.6 million adults, equivalent to 5% of adults). However, about 80% of long-term lonely people were affected by lockdown loneliness during the pandemic [3].

The increase in loneliness during the COVID-19 pandemic has been attributed to increased social isolation because of lockdowns, social distancing, self-isolation, and quarantine measures aimed at reducing the spread of coronavirus [12]. The COVID-19 pandemic is therefore being labelled as the pandemic of loneliness [13,14].

People More Likely to Be Affected by Lockdown Loneliness

The COVID-19 pandemic has not only resulted in disease-related illness and deaths but also has had serious adverse economic and sociopsychological impacts [13]. Lockdowns and social distancing during the COVID-19 pandemic have increased the risk of loneliness [11]. Recent studies showed that loneliness due to COVID-19 lockdowns, which we consider "lockdown loneliness," is higher in adults who are single, divorced, separated, widowed, and/or living alone, as well as those individuals who have bad to very bad health [3]. In addition, lockdown loneliness has increased in young people (aged 16-25 years) [3,11] and seniors (>70 years old) [14]; however, older adults (55-69 years old) are reported to be less likely to be affected by lockdown loneliness [3].

In addition, the risk of loneliness in people of ethnic minority background has increased during COVID-19 lockdowns [12]. Fancourt et al [15] reported 35% higher loneliness in people of Black, Asian, and minority ethnic (BAME) origin compared to white British people (23% BAME versus 17% White), while a study by the British Red Cross reported about 12% higher prevalence of loneliness in BAME people compared to white people (46% BAME versus 41% White) during lockdowns [12]. Moreover, groups who were less likely to be affected by loneliness prior to the COVID-19 pandemic, such as families with young children, have also been affected by lockdown loneliness during the pandemic [12], but there is no statistically significant difference in lockdown loneliness and prelockdown loneliness in this group [3].

Risk of lockdown loneliness during the pandemic may be greater in people with limitations such as hearing loss [16], people who are digitally excluded [12], and those who are disconnected from colleagues because of working from home, which has been identified as a risk factor for loneliness [17].

Factors Contributing to Lockdown Loneliness

The increase in loneliness during the pandemic has been attributed to COVID-19 lockdowns [3,11], during which social connection and social support become very limited because of the enforced physical distancing, social isolation, and quarantine measures [18]. These preventative measures have removed access to typical places used for social connection, interaction, and support [19], resulting in loneliness that could adversely affect physical and mental health and well-being [20]. In addition, COVID-19–related social distancing and isolation could result in sociopsychological harm, increasing the risk of loneliness in the most vulnerable and high-risk individuals, especially those who are socially, psychologically, and economically disadvantaged [19].

Loneliness is seldom observed in people who have social interaction, and socially active people have better overall health compared to those individuals who do not interact socially with others [21]. Loneliness is a social determinant of health [9] that is more commonly prevalent in people living in large cities [12]

XSL•FO

RenderX

and areas that are deprived and geographically remote [22]. Other risk factors for loneliness include personal circumstances and characteristics, health and disability, and life transitions [23]. A higher risk of lockdown loneliness has been reported in females, younger people, and people who are dissatisfied with family, have negative self-perceptions about aging, have less contact with relatives, have the self-perception of being a burden on family and friends for support, listen to news related to COVID-19, have fewer resources for self-entertaining, and are digitally excluded [12,24].

Lockdown Loneliness as an Important Public Health Issue

Loneliness is a major public health issue because it is associated with increased morbidity and mortality [4,5]. Loneliness is one of the key challenges that must be dealt with during the COVID-19 pandemic [25]. The situation could become more serious because levels of loneliness could rise due to an increase in the number of sociopsychological and mental health cases in the aftermath of the pandemic [18]. Empirical evidence shows that quarantine and lockdowns during viral infection epidemics, such as the SARS epidemic, result in more annoyance, fear, frustration, helplessness, isolation, loneliness, nervousness, sadness, and worry, and less happiness [26]. Similarly, the COVID-19 outbreak has resulted in psychological stressors related to the longer duration of quarantine, fear of infection, anxiety, feeling helpless, frustration, boredom, insufficient supplies, inadequate information, financial loss, and stigma, which further increase social isolation and loneliness [12,20,27]. At the same time, mental health and affective response to COVID-19's threat to health are significantly associated with loneliness [28]. Moreover, the limited access to health care; social support (both formal and informal), interaction, and communication; economic, employment, and leisure opportunities; and other activities during the COVID-19 crisis has accelerated the risk of severe morbidity and mortality in high-risk individuals [12,29].

Tackling Lockdown Loneliness

Tackling the rising tide of loneliness requires strengthening social connections and supporting people affected by lockdown loneliness during the COVID-19 crisis [18]. This requires efforts aimed at mitigating social isolation and facilitating social connectedness [30]. For tackling social isolation and loneliness during the COVID-19 pandemic, the World Health Organization has recommended maintaining social networks and staying connected with family, friends, colleagues, and community members via digital means [31]. More importantly, digital technology has become vital for addressing loneliness during the pandemic because other means of addressing loneliness (such as social prescribing) have become difficult if not impossible to access during the lockdowns. Even social prescribing for tackling loneliness has become digital social prescribing because it requires the use of digital technology during the pandemic [32].

Digital Technology Tools for Social Connection During the COVID-19 Crisis

Digital technology is already a main feature of health systems and health and social care delivery [33], but its application has become critical during the COVID-19 pandemic [34]. Digital technology is enabling not only online and remote health consultations and a myriad of business activities but also connecting socially distant people during lockdowns and social distancing [35]. For example, digital technology enables online meetings, conferences, boardroom and team meetings, working from home [36], online teaching and learning, and even virtual cabinet meetings, which have all become almost the norm of daily life and business during the pandemic. Many technological companies, whether tech giants or start-ups, have either updated their existing portfolio of tools or developed new tools to fill the gap created by social distancing and lockdowns. A few examples of widely used online digital tools for social connection and networking include Zoom, Microsoft Teams, GoToMeetings, and Google Hangouts [37]. These tools are being used in developed [38] and developing countries [39] and have become more acceptable and widely adopted during the current pandemic. In addition, these virtual technologies are increasingly being used for providing social and cognitive support, supporting learning and teaching, enabling buying and selling, facilitating leisure and hobbies, and doing collaborative innovative research that can be done at a distance. It is expected that the use of these tools will increase and become a part of daily business in many fields, including health care, for a range of activities (eg, online medical consultations, treatment approaches, and interventions) [40].

More importantly, some companies are creating new products to help reduce loneliness and its impacts [41]. Two examples are the Spill online messaging app (an online mental health therapy platform) [42] and QuarantineChat, a one-on-one voice chat service that was developed to help people who are isolated during viral epidemics and emergencies beat boredom [43]. In addition, there are numerous other apps (eg, Headspace, Happify, and MindShift) that were developed to address mental health issues in general but could also be helpful in alleviating social isolation and loneliness during COVID-19 lockdowns [44].

Assessment of Digital Technology Tools for Social Connectedness and Users' Needs

A recent systematic review showed that a variety of digital tools, such as social media platforms, video conferencing, online voice and video networks, and social internet-based activities, are used for tackling loneliness in various settings [45]. These digital technology tools could be helpful in addressing lockdown loneliness during the COVID-19 pandemic. However, these tools must be assessed not only for their advantages but also for their limitations, including any negative impacts they may have on social relations, as the use of digital social media tools has been allegedly associated with the breakup of relationships

and domestic abuse and violence in some families during quarantine and social isolation amid the pandemic [46,47]. It is also essential to evaluate how digital technology companies collect, manage, and use user data and whether there are any issues with regard to personal data security, privacy, and safety [48]. These issues are very important, especially for people who are more vulnerable, such as people with cancer [49], neurological conditions, and mental health problems, who could suffer more from the adverse impacts of the pandemic [31].

Therefore, digital technology tools used for health issues including lockdown loneliness must be safe, effective, and evidence-based [50]. The Anxiety and Depression Association of America has assessed and rated a number of apps from the end users' perspective, including criteria such as effectiveness and ease of use [44]. Although most of these applications are for addressing mental health issues, they might be helpful in alleviating social isolation, loneliness, and mental health issues during COVID-19 lockdowns. However, before their adoption, these tools must be assessed for their accessibility, affordability, and acceptance by end users and patients [51].

Access to and Use of Digital Technology for Combating Lockdown Loneliness

Combating lockdown loneliness during the COVID-19 pandemic requires changing the ways we connect socially [52], often through reliable, secure, easy to use, and effective digital technology tools [53]. More importantly, people who are most vulnerable to the adverse effects of the COVID-19 pandemic must not be digitally excluded [12]; rather, they should be actively provided access to digital technology [54]. Some people (eg, older adults) might have a low level of technological

knowledge and literacy [55] and may therefore encounter difficulties and be less confident when using online digital technology tools [14]. Such people should be supported in developing their skills to effectively use these tools for social connection [12] to help alleviate lockdown loneliness during the pandemic. However, it is essential to consider and plan for the resolution of some other pertinent issues (eg, addressing the digital infrastructure [55], systems, and processes that may require development, upgrading, and investment) to support digital technology tools. The costs and maintenance of digital technological tools, and support and coordinated involvement of all key stakeholders (family and friends, social care providers, and clinicians and health allied professionals) are critical factors that must be taken into account to tackle lockdown loneliness.

Conclusions

Digital technology has undoubtedly become critical for reducing and preventing social, physical, and psychological risks during the COVID-19 pandemic and addressing the short- and long-term impacts of social isolation and lockdown loneliness [18]. Nonetheless, most people affected by social isolation and lockdown loneliness during the pandemic might not feel lonely yet because these effects may take some time to show up [56]. It is therefore imperative that digital technology should not only provide tools to improve social connectedness and help in reducing lockdown loneliness but also enable people at risk of loneliness to take measures to avoid social isolation during the COVID-19 pandemic and in its aftermath. However, access to and costs and knowledge of digital technology tools are among the key issues that need urgent attention. Finally, tackling lockdown loneliness will require the active involvement of all key stakeholders that use these digital technology tools.

Authors' Contributions

All authors were involved in the planning, conception, and design of the study. SGSS drafted the manuscript. DN, VK, and HCvW reviewed the manuscript for intellectual input. VK helped in the acquisition of funds for paying open access publication charges. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Johns Hopkins Coronavirus Resource Center. COVID-19 Dashboard Internet. URL: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-09-15]
2. Farooq A, Laato S, Islam AKMN. Impact of Online Information on Self-Isolation Intention During the COVID-19 Pandemic: Cross-Sectional Study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
3. Office for National Statistics. Coronavirus and loneliness, Great Britain. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/coronavirusandlonelinessgreatbritain/3aprilto3may2020 [accessed 2020-06-11]

XSL•FO

RenderX

4. Rico-Uribe LA, Caballero FF, Martín-María N, Cabello M, Ayuso-Mateos JL, Miret M. Association of loneliness with all-cause mortality: A meta-analysis. PLoS One 2018 Jan 4;13(1):e0190033 [FREE Full text] [doi: 10.1371/journal.pone.0190033] [Medline: 29300743]

5. Fakoya OA, McCorry NK, Donnelly M. Loneliness and social isolation interventions for older adults: a scoping review of reviews. BMC Public Health 2020 Mar 14;20(1):129 [FREE Full text] [doi: 10.1186/s12889-020-8251-6] [Medline: 32054474]

6. Cacioppo JT, Cacioppo S, Boomsma DI. Evolutionary mechanisms for loneliness. Cogn Emot 2014;28(1):3-21 [FREE Full text] [doi: 10.1080/02699931.2013.837379] [Medline: 24067110]

7. HM Government. A connected society. A strategy for tackling loneliness - laying the foundations for change. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/750909/6. 4882_DCMS_Loneliness_Strategy_web_Update.pdf [accessed 2020-09-11]

8. VanderWeele TJ, Hawkley LC, Cacioppo JT. On the reciprocal association between loneliness and subjective well-being. Am J Epidemiol 2012 Nov 01;176(9):777-784 [FREE Full text] [doi: 10.1093/aje/kws173] [Medline: 23077285]

9. Murthy V. Work and the loneliness epidemic. Harvard Business Review. 2017. URL: https://hbr.org/cover-story/2017/09/work-and-the-loneliness-epidemic [accessed 2020-05-12]

10. Shah SGS, Nogueras D, van Woerden H, Kiparoglou V. Effectiveness of digital technology interventions to reduce loneliness in adults: a protocol for a systematic review and meta-analysis. BMJ Open 2019 Sep 27;9(9):e032455. [doi: 10.1136/bmjopen-2019-032455] [Medline: 31562164]

11. Mental Health Foundation. Loneliness during coronavirus. URL: https://www.mentalhealth.org.uk/coronavirus/coping-with-loneliness [accessed 2020-05-21]

12. British Red Cross. Life after lockdown: Tackling loneliness among those left behind. URL: https://www.redcross.org.uk/-/media/documents/about-us/research-publications/health-and-social-care/life-after-lockdown-tackling-loneliness-among-those-left-behind-report.pdf [accessed 2020-06-19]

13. Shah SGS, Farrow A. A commentary on "World Health Organization declares global emergency: A review of the 2019 novel Coronavirus (COVID-19)". Int J Surg 2020 Apr;76:128-129 [FREE Full text] [doi: 10.1016/j.ijsu.2020.03.001] [Medline: 32169574]

14. Live-in care agency finds 1 in 3 elderly people more lonely in wake of COVID-19. The Elder Magazine. URL: https://www.elder.org/the-elder/survey-on-elderly-loneliness/ [accessed 2020-09-14]

15. Fancourt D, Bu F, Mak H, Steptoe A. Covid-19 Social Study. Results Release 15. URL: https://www.covidsocialstudy.org/results [accessed 2020-08-13]

16. Tagupa H. Social Isolation, Loneliness, and Hearing Loss During COVID-19. The Hearing Journal 2020;73(5):46. [doi: 10.1097/01.HJ.0000666456.65020.b9]

17. Bevan S, Mason B, Bajorek Z. IES Working at Home Wellbeing Survey - Interim Findings. Institute for Employment Studies. URL: https://www.employment-studies.co.uk/resource/ies-working-home-wellbeing-survey [accessed 2020-08-13]

18. Sharma RA, Maheshwari S, Bronsther R. COVID-19 in the era of loneliness. Curr Psychiatry 2020;19(5):31-33 [FREE Full text]

19. Razai MS, Oakeshott P, Kankam H, Galea S, Stokes-Lampard H. Mitigating the psychological effects of social isolation during the covid-19 pandemic. BMJ 2020 May 21;369:m1904. [doi: 10.1136/bmj.m1904] [Medline: 32439691]

20. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]

21. Loneliness and the Workplace U.S. report. Cigna. 2020. URL: https://www.cigna.com/static/www-cigna-com/docs/about-us/newsroom/studies-and-reports/combatting-loneliness/cigna-2020-loneliness-report.pdf [accessed 2020-05-21]

22. Bolton M. Loneliness - the state we're in. 2012. URL: https://www.campaigntoendloneliness.org/wp-content/uploads/Loneliness-The-State-Were-In.pdf [accessed 2020-06-19]

23. Campaign to End Loneliness. Fact Sheet: Risk Factors. 2015. URL: https://campaigntoendloneliness.org/guidance/wp-content/uploads/2015/06/Risk-factorsGFLA.pdf [accessed 2020-05-21]

24. Losada-Baltar A, Jiménez-Gonzalo L, Gallego-Alberto L, Pedroso-Chaparro MDS, Fernandes-Pires J, Márquez-González M. "We're staying at home". Association of self-perceptions of aging, personal and family resources and loneliness with psychological distress during the lock-down period of COVID-19. J Gerontol B Psychol Sci Soc Sci 2020 Apr 13:1 [FREE Full text] [doi: 10.1093/geronb/gbaa048] [Medline: 32282920]

25. Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015 May;105(5):1013-1019. [doi: 10.2105/AJPH.2014.302427] [Medline: 25790413]

26. Reynolds DL, Garay JR, Deamond SL, Moran MK, Gold W, Styra R. Understanding, compliance and psychological impact of the SARS quarantine experience. Epidemiol Infect 2008 Jul;136(7):997-1007. [doi: 10.1017/S0950268807009156] [Medline: 17662167]

XSL•FO

RenderX

27. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]

28. Okruszek L, Aniszewska-Stańczuk A, Piejka A, Wiśniewska M, Żurek K. Safe but lonely? Loneliness, mental health symptoms and COVID-19. PsyArXiv. Preprint posted online on April 4, 2020. [doi: 10.31234/osf.io/9njps]

29. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ 2020 Mar 26;368:m1198. [doi: 10.1136/bmj.m1198] [Medline: 32217618]

30. Finley D. How increased social distancing for the coronavirus could spur a loneliness epidemic. Business Insider. 2020. URL: https://www.businessinsider.com/coronavirus-could-spur-loneliness-epidemic-2020-3 [accessed 2020-04-10]

31. World Health Organization. Mental health and psychosocial considerations during the COVID-19 outbreak. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf [accessed 2020-06-04]

32. Martin Bell Partnership and Associates. Social Prescribing and the Digital Landscape. 2018. URL: https://www.london.gov.uk/what-we-do/health/social-prescribing/social-prescribing-resources [accessed 2020-08-17]

33. World Health Organization. WHO guideline recommendations on digital interventions for health system strengthening. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-06-04]

34. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]

35. Manjunath BS. Covid-19: 8 ways in which technology helps pandemic management. ET CIO. 2020 Apr 14. URL: https://cio.economictimes.indiatimes.com/news/next-gen-technologies/covid-19-8-ways-in-which-technology-helps-pandemic-management/75139759 [accessed 2020-05-21]

36. Osborne S. Why remote working doesn't have to mean alienated employees. World Economic Forum. URL: https://www.weforum.org/agenda/2020/04/remote-working-workforce-feedback/ [accessed 2020-06-05]

37. Trueman C. Pandemic leads to surge in video conferencing app downloads. Computer World. URL: https://www.computerworld.com/article/3535800/pandemic-leads-to-surge-in-video-conferencing-app-downloads.html [accessed 2020-06-05]

38. Krausz M, Westenberg JN, Vigo D, Spence RT, Ramsey D. Emergency Response to COVID-19 in Canada: Platform Development and Implementation for eHealth in Crisis Management. JMIR Public Health Surveill 2020 May 15;6(2):e18995 [FREE Full text] [doi: 10.2196/18995] [Medline: 32401218]

39. Kadir MA. Role of telemedicine in healthcare during COVID-19 pandemic in developing countries. TMT 2020 Apr 30;5(2):187. [doi: 10.30953/tmt.v5.187]

40. Webster P. Virtual health care in the era of COVID-19. The Lancet 2020 Apr;395(10231):1180-1181. [doi: 10.1016/s0140-6736(20)30818-7]

41. Gilchrist K. Loneliness is on the rise due to the coronavirus: These entrepreneurs are making tackling isolation their business. CNBC. URL: https://www.cnbc.com/2020/05/13/entrepreneurs-startups-are-tackling-coronavirus-isolation-loneliness.html [accessed 2020-06-05]

42. Spill App Ltd. Spill chat. URL: https://www.spill.chat/ [accessed 2020-06-05]

43. Baskin D, Hawkins M. QuarantineChat. URL: https://quarantinechat.com/ [accessed 2020-06-05]

44. Anxiety and Depression Association of America. ADAA Reviewed Mental Health Apps. URL: https://adaa.org/finding-help/mobile-apps [accessed 2020-08-18]

45. Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V. Are digital technology interventions effective to reduce loneliness in older adults? A systematic review and meta-analysis. medRxiv. Preprint posted online on September 1, 2020. [doi: 10.1101/2020.08.27.20183012]

46. Leonard E. Breakup, Separation, and Divorce During COVID-19: Unprecedented times call for unprecedented support. Psychology Today. URL: https://www.psychologytoday.com/us/blog/peaceful-parenting/202004/breakup-separation-and-divorce-during-covid-19 [accessed 2020-06-05]

47. Douglas M, Katikireddi SV, Taulbut M, McKee M, McCartney G. Mitigating the wider health effects of covid-19 pandemic response. BMJ 2020 Apr 27;369:m1557. [doi: 10.1136/bmj.m1557] [Medline: 32341002]

48. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]

49. Meti N, Rossos PG, Cheung MC, Singh S. Virtual Cancer Care During and Beyond the COVID-19 Pandemic: We Need to Get It Right. JCO Oncology Practice 2020 Sep 13;16(9):527-529. [doi: 10.1200/op.20.00281]

50. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly 2018;148:w14571 [FREE Full text] [doi: 10.4414/smw.2018.14571] [Medline: 29376547]

51. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019 Aug 30;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]

52. Pallarés Carratalá V, Górriz-Zambrano C, Llisterri Caro J, Gorriz J. [The COVID-19 pandemic: An opportunity to change the way we care for our patients]. Semergen 2020 Aug;46 Suppl 1:3-5 [FREE Full text] [doi: 10.1016/j.semerg.2020.05.002] [Medline: 32446589]

53. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242. [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]

54. Local Government Association. Loneliness, social isolation and COVID-19 Internet. URL: https://tinyurl.com/yxpa76g8 [accessed 2020-06-04]

55. Conroy KM, Krishnan S, Mittelstaedt S, Patel SS. Technological advancements to address elderly loneliness: practical considerations and community resilience implications for COVID-19 pandemic. WWOP 2020 Sep 08;ahead-of-print(ahead-of-print):1. [doi: 10.1108/wwop-07-2020-0036]

56. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]

Abbreviations

BAME: Black, Asian, and minority ethnic

XSL•FO

RenderX

Viewpoint

The Internet of Things: Impact and Implications for Health Care Delivery

Jaimon T Kelly[1,2], BHlthSc, Mast Nutr&Diet, PhD; Katrina L Campbell[1,2,3], PhD; Enying Gong[4], MSc; Paul Scuffham[1,2], PhD

[1]Menzies Health Institute Queensland, Griffith University, Gold Coast, Australia

[2]Centre of Applied Health Economics, Griffith University, Brisbane, Australia

[3]Metro North Hospital and Health Service, Brisbane, Australia

[4]School of Population and Global Health, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Jaimon T Kelly, BHlthSc, Mast Nutr&Diet, PhD
Menzies Health Institute Queensland
Griffith University
G40 Griffith Health Centre
Level 8.86 Gold Coast campus
Gold Coast, 4222
Australia
Phone: 61 5678001 ext 8001
Email: jaimon.kelly@griffith.edu.au

Abstract

The Internet of Things (IoT) is a system of wireless, interrelated, and connected digital devices that can collect, send, and store data over a network without requiring human-to-human or human-to-computer interaction. The IoT promises many benefits to streamlining and enhancing health care delivery to proactively predict health issues and diagnose, treat, and monitor patients both in and out of the hospital. Worldwide, government leaders and decision makers are implementing policies to deliver health care services using technology and more so in response to the novel COVID-19 pandemic. It is now becoming increasingly important to understand how established and emerging IoT technologies can support health systems to deliver safe and effective care. The aim of this viewpoint paper is to provide an overview of the current IoT technology in health care, outline how IoT devices are improving health service delivery, and outline how IoT technology can affect and disrupt global health care in the next decade. The potential of IoT-based health care is expanded upon to theorize how IoT can improve the accessibility of preventative public health services and transition our current secondary and tertiary health care to be a more proactive, continuous, and coordinated system. Finally, this paper will deal with the potential issues that IoT-based health care generates, barriers to market adoption from health care professionals and patients alike, confidence and acceptability, privacy and security, interoperability, standardization and remuneration, data storage, and control and ownership. Corresponding enablers of IoT in current health care will rely on policy support, cybersecurity-focused guidelines, careful strategic planning, and transparent policies within health care organizations. IoT-based health care has great potential to improve the efficiency of the health system and improve population health.

Introduction

The challenges presented by an aging population with multiple chronic conditions are ubiquitous worldwide [1]. The medical, lifestyle, and personal health needs across aging populations will continue to place a burden on health care resources. Meeting these challenges requires a focus on empowering populations to self-manage their health through health innovation to improve well-being and attenuate health resource burden [2].

Background of Digital Devices and the Internet of Things

Entering the 2020 decade, more devices are connected to the internet than ever before, and this will continue to grow at a

rapid trajectory. Worldwide, more than 21 billion devices have been estimated to be connected to the internet in 2020, which is 5 times the number of devices 4 years prior [3]. The Internet of Things (IoT) can be defined in its simplest scenario as a network that connects uniquely identifiable devices (or *things*) to the internet, enabling them to collect, send, store, and receive data [4]. From a health care perspective, IoT can be considered as any device that can collect health-related data from individuals, including computing devices, mobile phones, smart bands and wearables, digital medications, implantable surgical devices, or other portable devices, which can measure health data and connect to the internet [5].

The growth of IoT technology has driven interest in a wide range of health practices to improve population health more specifically [6]. Recent reviews have overviewed the various services and applications of IoT in health care (eg, eHealth, mobile health [mHealth], ambient assisted living, semantic devices, wearable devices and smartphones, and community-based health care) [5,7]. These services have been detailed extensively and can have many applications across single condition and cluster condition management, including, for example, the ability to track and monitor health progress remotely by health care professionals, improve self-management of chronic conditions, assist in the early detection of abnormalities, fast-track symptom identification and clinical diagnoses, deliver early intervention, and improve adherence to prescriptions [8]. These applications can make better use of health care resources and provide quality and low-cost medical care.

Health Systems Are Changing

With the 2020 public health response to the novel COVID-19 pandemic to effectively shut down traditional modes of health service delivery worldwide, efforts to reduce implementation barriers to technology-supported health delivery highlight the potential to reframe traditional models of care into virtual and distance modalities [9]. In response, many countries have successfully implemented technology-supported services to maintain health care practices and social distancing [10]. As global leaders consider policies that potentially provide more access to technology-supported health services in response to (and considerations post) the current COVID-19 crisis, it is becoming increasingly important to understand how established and emerging IoT technologies can support health systems to deliver safe and effective care in either a complementary or an alternative way during times of crisis or health epidemics [11].

This viewpoint paper will overview current technologies in health care, outline how IoT devices are improving health service delivery, and outline how IoT technologies can affect global health care in the next decade. This viewpoint paper also overviews how the disruption in health care from IoT can lead to improved access and equitable primary, secondary, and tertiary smart health care, which is more proactive, continuous, and coordinated.

IoT-Based Health Care Architecture

The architecture of IoT in health care delivery essentially consists of 3 basic layers [12]: (1) the perception layer, (2) the network layer, and (3) the application layer. It is not our intention to extensively detail these layers; however, a summary and the related health implications are provided in the following sections.

Perception Layer: Sensing Systems That Collect Data

Perception and identification technologies are the foundation of IoT. Sensors are devices that can perceive changes in an environment and can include, for example, radio frequency identification (RFID), infrared sensors, cameras, GPS, medical sensors, and smart device sensors. These sensors allow for comprehensive perception through object recognition, location recognition, and geographic recognition and can convert this information to digital signals, which is more convenient for network transmission [12,13]. Sensor technologies allow for treatments to be monitored in real time and facilitate the acquisition of a multitude of physiological parameters about a patient so that diagnoses and high-quality treatment can be fast-tracked. There are many examples of potentially lifesaving IoT sensor devices; however, not all devices are clinically tested or have been proved to be safe or effective. A summary of IoT devices that may support and improve health service delivery is provided in Multimedia Appendix 1 [14-47].

Network Layer: Data Communication and Storage

The network level of IoT technologies includes wired and wireless networks, which communicate and store processed (layer 1) information either locally or at a centralized location. Communication between things can occur over low, medium, and high frequencies, the latter being the predominant focus of IoT. These include short-range communication technologies, such as RFID, wireless sensor networks, Bluetooth, Zigbee, low-power Wi-Fi, and global system for mobile communications [12]. High-frequency fourth-generation (4G) cellular networks have seen even more communication potential, and evolving 5G networks are becoming more readily available and are expected to be a major driver of the growth of IoT applications for health care, with the potential to provide reliable connection up to thousands of devices at the same time [48].

Communicated data are stored locally (often decentralized) or sent to a centralized cloud server. Cloud-based computing to support the delivery of health services has many benefits, as it is ubiquitous, flexible, and scalable in terms of data acquisition, storage, and transmission between devices connected to the cloud [49]. The use of the cloud can be foreseen to support data-intensive electronic medical records (EMRs), patient portals, medical IoT devices (which can include smartphone apps), and the big data analytics driving decision support systems and therapeutic strategies [5]. However, with more cloud apps entering the health market, it is just as important that an evidence base supports its effectiveness and safety and can deal with the security of health data and the reliability and transparency of that data by third parties. Furthermore, it has been suggested that centralized cloud storage will present issues in the future to users, such as excessive data accumulation and

latency because of the distance between IoT devices and data centers.

Decentralized data processing and networking approaches may improve the scalability of IoT in health care. Edge cloud is a newer cloud computing concept that allows IoT sensors and network gateways to process and analyze data themselves (ie, at the *edge*) in a decentralized fashion, reducing the amount of data required to be communicated and managed at a centralized location [12,50]. Similarly, blockchain storage uses a decentralized approach to data storage, creating independent blocks containing individual sets of information, which forms a dependent link in a collective block, which in turn creates a network regulated by patients rather than a third party [51]. There are examples of platforms engineering blockchain for medical practice already [51,52]; however, research on edge cloud and blockchains in health care is still limited and is an important area for future research.

Application Layer

The application layer interprets and applies data and is responsible for delivering application-specific services to the user [12]. Some of the most promising medical applications that IoT provides are through artificial intelligence (AI). The scientific applications of AI have proliferated, including image analysis, text recognition with natural language processing, drug activity design, and prediction of gene mutation expression [53]. AI has the capability to read available EMR data, including medical history, physical, laboratory, imaging, and medications, and contextualize these data to generate treatment and/or diagnosis decisions and/or possibilities. For example, IBM Watson uses AI to read both structured and unstructured text in the EMR, read images to highlight primary and incidental findings, and compile relevant medical literature in response to clinical queries [54].

IoT-based health care and use of deep machine learning can assist health professionals in seeing the unseeable and providing new and enhanced diagnostic capability. Although diagnostic confidence may never reach 100%, combining machines and clinician expertise reliably enhances system performance. For example, compared with the diagnostic evaluation by 54 ophthalmologists and senior residents, applying AI to retinal images improved the detection and grading of diabetic retinopathy and macular edema, achieving high specificities (98%) and sensitivities (90%) [55]. AI and deep learning can also optimize disease management, can provide big data and analysis generated from mHealth apps and IoT devices, and are starting to see adoption in health care [56]. Some examples of this include predicting risk, future medical outcomes, and care decisions in diabetes and mental health [57] and predicting the progression of congestive heart failure [58,59], bone disease [60], Alzheimer disease [61], benign and malignant tumor classification [62,63], and cardiac arrhythmias [64].

Expanding the Functions and Scope of IoT to Provide Smart Health Care

IoT is an infrastructure that enables smart health services to operate. When health data are collected by IoT sensors, communicated, and stored, this enables data analytics and smart health care, which can improve risk factor identification, disease diagnoses, treatment, and remote monitoring and empower people to self-manage.

Smart health care services make use of advancements in information technologies, such as IoT, big data analytics, cloud computing, AI, and deep machine learning, to transform traditional health care delivery to be a more efficient, convenient, and a more personalized system [65]. Current developments in information computer technologies have allowed the development of health care solutions with more intelligent prediction capabilities both in and out of the hospital. We are seeing the use of virtual models to transfer care provided in hospitals to the home through the use of sensors and devices that allow remote review and monitoring of patients in their homes or treated in hospitals and creates a continuum among these through cloud access [7]. More recently, the 2020 public health efforts around the world to mitigate the spread of COVID-19 have (at least temporarily) led governments and policy makers to remove implementation and remuneration barriers to enable health care professionals to use virtual models of care for people who need it [9]. IoT also provides the opportunity to improve the quality and efficiency of the entire ecosystem of service delivery, including hospital management, medical asset management, monitoring of the workflow of staff, and optimization of medical resources based on patient flow [66,67].

How IoT Can Improve Health Service Delivery

Primary Health Care Becoming More Accessible

A focus on disease prevention must become a priority this decade, as the burden of disease attributable to modifiable risk factors is greater than ever before [1,68]. IoT in health care has the potential to improve population health and transition our health care model to a true hybrid model of primary, secondary, and tertiary care, where the health system can use its existing workforce in new and more efficient ways. Transforming health delivery in this way is crucial to improving self-management for people with chronic conditions, as even among high health care users, more than 90% of lifestyle self-management is done by patients themselves, outside of hospitals, and in clinical settings [69,70].

There is a clear public demand for easy-to-access health information. For example, in a 2015 US survey, 58% (931/1604) of smartphone users downloaded a health-related app for their lifestyle self-management [71]. AI has also driven the availability of point-of-care health information, such as chatbots (or AI doctors), which can deliver lifestyle and medical advice. Examples of these established AI *bots* are Woebot, Your.Md, Babylon, and HealthTap, where a patient can input their symptoms and advice is generated instantly [72]. However, more than half of the most highly rated apps make medical claims that are not approved [73], with no formal process of approving apps or informing consumer choice [74], and much remains to be done to understand the potential of chatbots to improve health. Therefore, a reliable digital health evidence

base is essential [75]. If health professionals have evidence-based digital resources, devices, and mobile apps readily at their disposal, digital prescriptions could become an enabler of wider adoption of IoT in health care and facilitate a wider population focus on disease prevention.

At the individual level, IoT offers the opportunity to link and potentially learn from nonhealth IoT technologies to monitor daily activities, provide support with information, and promote behavior changes (Multimedia Appendix 2). In addition, IoT and data linkage create great potential of transparent, evidence-based decision making, which may be able to drive the shift of disease patterns and increase the well-being of citizens at scale. The integration of urban infrastructures, IoT technologies, and cloud computing allows the collection and analysis of a vast quantity of different human and non–human-related data. These data could provide valuable information about population-level surveillance in diseases and accidents, risk factors, and environmental conditions [76], which is difficult to collect through the traditional human-reported disease surveillance system and can be of particular benefit in pandemic responses [77]. For example, in Taiwan, big data analytics applied to electronic data (GPS, closed-circuit television surveillance, and credit card payments) in the community and personal mobile data have been effectively used to contact trace, communicate, and isolate potential contacts during the global COVID-19 pandemic [78]. Through IoT and data linkage, decision makers are likely to be able to make evidence-based decisions in promoting healthy social and built environments, safe transportation systems, high-quality public services, and smart health care and emergency response systems [76,79,80].

Secondary and Tertiary Health Care That Is Proactive, Continuous, and Coordinated

An IoT-based health care system enables the overall health care systems to move past a traditional model of service delivery, which is often reactive, intermittent, and uncoordinated, to a more proactive, continuous, and coordinated approach [81]. Such an approach is favorable because it offers the opportunity to provide high-quality care that is less invasive and appealing to patients and health care professionals. This change in the health care system landscape is also highly appealing for policy makers because it can greatly enhance the efficiency (and subsequently reduce resource use) of the health system [82] and also provide the health system flexibility to shift its models of care and delivery of services as required on an individual or population-wide basis. Multimedia Appendix 3 summarizes 7 examples of how IoT can improve the coordination of health services and likely improve our health system efficiency.

Enablers and Barriers to Address for IoT-Based Health Care

Enablers

Policy Support

Policy support is one of the most important environmental enablers of IoT. Many countries already have policies in place

for eHealth (eg, web-based and software programs to deliver health services) [83,84] and either have or are in the process of developing policies for IoT infrastructure, investment, and/or implementation in health care. For example, China, India, Indonesia, Japan, Malaysia, the Philippines, Singapore, Thailand, the European Union, the United States, and Vietnam currently have relevant policies in place for IoT [85]. Australia is also in the process of establishing a policy for IoT development and investment [86].

Technology That Is Accessible and Easy to Use

The ubiquitous nature of technology means that consumers and health care professionals have greater access to digital resources than ever before [87]. However, it is also important for health systems to be aware of the inequities that may eventuate from the widespread implementation of IoT for health care, including individuals who may not be able to afford or access technology hardware or reliable internet services because of geographic location or financial disadvantage. Similarly, if individuals do not perceive the technology as *user friendly*, experience poor connections, or do not feel the initiative has been designed in consultation with them (both patients and health professionals), then this often results in frustration and reluctance to use such services [88,89].

Cybersecurity-Focused Guidelines for Robust and Resilient Market Adoption

Cyber risk is a major obstacle to the broad adoption of IoT [90]. The privacy of patients must be ensured to prevent unauthorized identification and tracking. From this perspective, the higher the level of autonomy and intelligence of the things, the more the challenges for the protection of identities and privacy.

Barriers

Confidence and Acceptability

There is a gap in public awareness and understanding of data safety in cloud-stored health information. This is of concern, as it is the single biggest threat to the adoption of IoT from a societal perspective. The premise of IoT is clear to society; however, what is not clear to people is the actual value that IoT delivers to them personally from a health care perspective [91,92]. The potential threat of breached confidentiality may never go away; however, the perceived value to consumers needs to outweigh these concerns to confidently engage with IoT-supported health infrastructure [90]. The confidence and acceptability of IoT by health care professionals are similarly important. There is a diverse range of factors that affect clinicians' acceptability of technology-supported programs, including the characteristics of the technology (eg, accuracy, compatibility with usual systems, and ease of use), individual's attitudes and knowledge (eg, familiarity and impact on professional security), external factors (eg, patient and health professional interaction), and organization readiness (eg, training and reimbursement) [93].

Privacy and Security

IoT might allow opportunities for cyberattacks and for personal data to be collected inappropriately. IoT-based applications are vulnerable to cyberattacks for 2 basic reasons: (1) most of the

communications are wireless, which makes eavesdropping very easy; and (2) most of the IoT components are characterized by low energy, and therefore, they can hardly implement complex schemes on their own to ensure security. The National Institute of Standards and Technology has recently released a draft security guide and recommendations for IoT devices, which will see an emphasis on data security in IoT devices [94]; however, whether such a guideline can or will be enforced across IoT health devices is unclear.

Data Storage, Control, and Ownership

To move forward in IoT-based health care, transparency and enforced codes of practice regarding where centralized cloud data are stored and who owns the data, needs to be considered For example, does the data host have viewing rights to someone's data and are these data completely controlled by individuals or are they never deleted from the cloud, despite a user's request? Another important consideration is the sharing of data across states or territories and internationally. Privacy, security, and confidentiality of data control and storage should be federally enforced, but international hosts and suppliers may not be required to follow any such code. Therefore, the use of these platforms requires strategic planning and transparent guidelines to develop and implement robust IoT-based health care policies and models of care.

Interoperability and Standardization Protocols

Issues around the interoperability and standardization of IoT and health care systems are a big threat to the wider adoption of IoT for health care systems. Lack of standardization threatens the development of IoT in the health setting context, as the industry and manufacturers are yet to reach a consensus regarding wireless communication protocols and standards for machine-to-machine communication. Without a unified, standardized, and interoperable system, the adoption of IoT into health care will be greatly hindered and is unlikely to have international reach [95]. Semantic interoperability in IoT is a necessary condition for big data techniques to support decision-making processes [96]. It is increasingly common for each new technology startup, device, or system manufacturer to define their own specific architecture, protocols, and data formats, which are unable to communicate with the health care environment unless they are appreciably redeveloped or adapted to interoperate with hospital IoT platforms [96]. This creates *Vertical Silos* [97], which demands the development of new features for granting interoperability between different systems. The future and full potential of IoT-enabled health care relies on addressing interoperability, of which some frameworks do exist [98]. Achieving interoperability across IoT platforms can provide a safer, more accessible, productive, and satisfying experience for clinicians and patients alike.

Remuneration

Finally, remuneration for technology-assisted health care has historically been challenging [99] and differs appreciably across different countries. This is likely to be even more complex for IoT-delivered health care, where reimbursement considerations have not been established (and this is unlikely until the abovementioned points are addressed). As international health systems establish robust policies and guidelines on cybersecurity and address the issues surrounding interoperability and standardization protocols, reimbursement and regulatory considerations across single-payer and multipayer systems should become a key priority to ensuring successful, effective, and cost-effective IoT health care models can be implemented in practice.

Conclusions

From this viewpoint, the potential of IoT is summarized as a growing area of research in health care. These developments provide a great opportunity for health care systems to proactively predict health issues and diagnose, treat, and monitor patients both in and out of the hospital. As the adoption of technology-supported health services increases to enable health systems to deliver flexible models of care, an increasing number of traditional health service delivery practices will be complemented or replaced through IoT. However, the implementation of IoT in health care will rely on a clear and robust code of practice for the management of data, privacy, confidentiality, and cybersecurity concerning the supply and use of IoT devices in health care. There are still important gaps for future research to address, which relate to the IoT technology itself, the health system, and the users of IoT technology. Specific future research on IoT technology needs to address how IoT devices can be designed with standardized protocols and interoperability with international and cross-state health systems. More research is also needed on the efficiency of blockchain storage compared with centralized cloud-based storage solutions in the context of IoT-supported health care delivery. From a health system perspective, there is a need for clinical guidelines on digital health prescriptions and robust policy regarding remuneration for primary and secondary care services provided through IoT. Finally, more research is needed to determine the acceptability and digital literacy of consumers and clinicians in the context of using IoT to improve the delivery and overall experience of health care. Although this viewpoint is a summary of selected literature only and not based on an exhaustive systematic review of the literature, we believe that addressing these areas for future research will go a long way to enable a wider uptake of IoT, which can ultimately save health care dollars and improve patient-centered care.

XSL·FO

RenderX

Authors' Contributions

Each author contributed to the conception and design of this paper. JK conducted the literature searches and drafted the first draft of the manuscript. EG, KC, and PS revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of Internet of Things devices that can support health service delivery.
[DOCX File , 20 KB - jmir_v22i11e20135_app1.docx]

Multimedia Appendix 2
Examples of how smart homes can improve health care delivery.
[DOCX File , 14 KB - jmir_v22i11e20135_app2.docx]

Multimedia Appendix 3
Scenarios where Internet of Things can be used to improve health system efficiency.
[DOCX File , 14 KB - jmir_v22i11e20135_app3.docx]

References

1.   GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]

2.   Mitchell M, Kan L. Digital technology and the future of health systems. Health Syst Reform 2019;5(2):113-120. [doi: 10.1080/23288604.2019.1583040] [Medline: 30908111]

3.   Internet of Things (IoT) in Healthcare. Research Markets. 2019. URL: https://www.medicaldevice-network.com/comment/bringing-internet-things-healthcare/ [accessed 2020-10-02]

4.   Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]

5.   Dang LM, Piran MJ, Han D, Min K, Moon H. A survey on internet of things and cloud computing for healthcare. Electronics 2019 Jul 9;8(7):768. [doi: 10.3390/electronics8070768]

6.   Saarikko T, Westergren UH, Blomquist T. The internet of things: are you ready for what's coming? Bus Horiz 2017 Sep;60(5):667-676. [doi: 10.1016/j.bushor.2017.05.010]

7.   Nazir S, Ali Y, Ullah N, García-Magariño I. Internet of things for healthcare using effects of mobile computing: a systematic literature review. Wireless Commun Mobile Comput 2019 Nov 14;2019:1-20. [doi: 10.1155/2019/5931315]

8.   Yin Y, Zeng Y, Chen X, Fan Y. The internet of things in healthcare: an overview. J Ind Inf Integration 2016 Mar;1:3-13. [doi: 10.1016/j.jii.2016.03.004]

9.   Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and covid-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]

10. Fisk M, Livingstone A, Pit SW. Telehealth in the context of covid-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 9;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]

11. Ye Q, Zhou J, Wu H. Using information technology to manage the covid-19 pandemic: development of a technical framework based on practical experience in China. JMIR Med Inform 2020 Jun 8;8(6):e19515 [FREE Full text] [doi: 10.2196/19515] [Medline: 32479411]

12. Sethi P, Sarangi S. Internet of things: architectures, protocols, and applications. J Electric Comput Eng 2017:-.

13. Wu M, Lu T, Ling F, Sun J. Research on the Architecture of Internet of Things. In: 3rd International Conference on Advanced Computer Theory and Engineering. 2010 Presented at: ICACTE'10; August 20-22, 2010; Chengdu, China. [doi: 10.1109/icacte.2010.5579493]

14. Merchant RK, Inamdar R, Quade RC. Effectiveness of population health management using the propeller health asthma platform: a randomized clinical trial. J Allergy Clin Immunol Pract 2016;4(3):455-463. [doi: 10.1016/j.jaip.2015.11.022] [Medline: 26778246]

15. Merchant R, Szefler SJ, Bender BG, Tuffli M, Barrett MA, Gondalia R, et al. Impact of a digital health intervention on asthma resource utilization. World Allergy Organ J 2018;11(1):28 [FREE Full text] [doi: 10.1186/s40413-018-0209-0] [Medline: 30524644]

XSL•FO
RenderX

16. Levine DM, Ouchi K, Blanchfield B, Diamond K, Licurse A, Pu CT, et al. Hospital-level care at home for acutely ill adults: a pilot randomized controlled trial. J Gen Intern Med 2018 May;33(5):729-736 [FREE Full text] [doi: 10.1007/s11606-018-4307-z] [Medline: 29411238]

17. Plowman R, Peters-Strickland T, Savage G. Digital medicines clinical review on the safety of tablets with sensors. 2018;17(9):849-852 [FREE Full text]

18. Naik R, Macey N, West RJ, Godbehere P, Thurston SC, Fox R, et al. First use of an ingestible sensor to manage uncontrolled blood pressure in primary practice: the UK hypertension registry. J Community Med Health Educ 2017;07(01):-. [doi: 10.4172/2161-0711.1000506]

19. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]

20. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]

21. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020 Jan 18;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]

22. Tashjian VC, Mosadeghi S, Howard AR, Lopez M, Dupuy T, Reid M, et al. Virtual reality for management of pain in hospitalized patients: results of a controlled trial. JMIR Ment Health 2017 Mar 29;4(1):e9 [FREE Full text] [doi: 10.2196/mental.7387] [Medline: 28356241]

23. Chirico A, Lucidi F, De Laurentiis M, Milanese C, Napoli A, Giordano A. Virtual reality in health system: beyond entertainment. A mini-review on the efficacy of VR during cancer treatment. J Cell Physiol 2016 Feb;231(2):275-287. [doi: 10.1002/jcp.25117] [Medline: 26238976]

24. de Carvalho M, Dias T, Duchesne M, Nardi A, Appolinario J. Virtual reality in health system: beyond entertainment. A mini-review on the efficacy of VR during cancer treatment. Behav Sci (Basel) 2017 Jul 9;7(3):- [FREE Full text] [doi: 10.3390/bs7030043] [Medline: 28698483]

25. Lohse KR, Hilderman CG, Cheung KL, Tatla S, van der Loos HF. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]

26. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]

27. Persky S. Application of virtual reality methods to obesity prevention and management research. J Diabetes Sci Technol 2011 Mar 1;5(2):333-339 [FREE Full text] [doi: 10.1177/193229681100500220] [Medline: 21527102]

28. Sangave NA, Aungst TD, Patel DK. Smart connected insulin pens, caps, and attachments: a review of the future of diabetes technology. Diabetes Spectr 2019 Nov;32(4):378-384. [doi: 10.2337/ds18-0069] [Medline: 31798296]

29. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and FitBit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]

30. Sposaro F, Tyson G. iFall: an Android application for fall monitoring and response. (1557-170X (Print)) 2020:a. [doi: 10.1109/iembs.2009.5334912]

31. Majumder S, Chen L, Marinov O, Chen C, Mondal T, Deen MJ. Noncontact wearable wireless ECG systems for long-term monitoring. IEEE Rev Biomed Eng 2018;11:306-321. [doi: 10.1109/rbme.2018.2840336]

32. Kakria P, Tripathi NK, Kitipawang P. A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. Int J Telemed Appl 2015;2015:373474 [FREE Full text] [doi: 10.1155/2015/373474] [Medline: 26788055]

33. Dojchinovski D, Ilievski A, Gusev M. editors. Interactive home healthcare system with integrated voice assistant 2019;2019:-. [doi: 10.23919/mipro.2019.8756983]

34. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]

35. Ilievski A, Dojchinovski D, Gusev M. Interactive Voice Assisted Home Healthcare Systems. In: Proceedings of the 9th Balkan Conference on Informatics. 2019 Presented at: BCI'19; September 26-28, 2019; University of Sofia, Bulgaria. [doi: 10.1145/3351556.3351572]

36. Jadczyk T, Kiwic O, Khandwalla RM, Grabowski K, Rudawski S, Magaczewski P, et al. Feasibility of a voice-enabled automated platform for medical data collection: CardioCube. Int J Med Inform 2019 Sep;129:388-393. [doi: 10.1016/j.ijmedinf.2019.07.001] [Medline: 31445282]

37. Chen S, Jones C, Moyle W. Social robots for depression in older adults: a systematic review. J Nurs Scholarsh 2018 Nov;50(6):612-622. [doi: 10.1111/jnu.12423] [Medline: 30198077]

38. van der Putte D, Boumans R, Neerincx M, Rikkert M, de MM. A Social Robot for Autonomous Health Data Acquisition Among Hospitalized Patients: An Exploratory Field Study. In: 14th ACM/IEEE International Conference on Human-Robot

Interaction (HRI). 2019 Presented at: HRI'19; March 11-14, 2019; Daegu, Korea (South), Korea (South). [doi: 10.1109/hri.2019.8673280]

39. Wan S, Gu Z, Ni Q. Cognitive computing and wireless communications on the edge for healthcare service robots. Comput Comm 2020 Jan;149:99-106. [doi: 10.1016/j.comcom.2019.10.012]

40. Baur K, Schättin A, de Bruin ED, Riener R, Duarte JE, Wolf P. Trends in robot-assisted and virtual reality-assisted neuromuscular therapy: a systematic review of health-related multiplayer games. J Neuroeng Rehabil 2018 Nov 19;15(1):107 [FREE Full text] [doi: 10.1186/s12984-018-0449-9] [Medline: 30454009]

41. Tao V, Moy K, Amirfar VA. A little robot with big promise may be future of personalized health care. Pharmacy Today 2016 Sep;22(9):38. [doi: 10.1016/j.ptdy.2016.08.022]

42. Ritschel H, Seiderer A, Janowski K, Aslan I, André E. Drink-O-Mender: An Adaptive Robotic Drink Adviser. In: Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction. 2018 Presented at: MHFI'18; October 16, 2018; Boulder, Colorado. [doi: 10.1145/3279954.3279957]

43. Moyle W, Jones C, Cooke M, O'Dwyer S, Sung B, Drummond S. Social Robots Helping People With Dementia: Assessing Efficacy of Social Robots in the Nursing Home Environment. In: 6th International Conference on Human System Interactions (HSI). 2013 Presented at: HSI'13; June 6-8, 2013; Sopot, Poland. [doi: 10.1109/hsi.2013.6577887]

44. Bhelonde A, Didolkar N, Jangale S, Kulkarni N. Flexible Wound Assessment System for Diabetic Patient Using Android Smartphone. In: International Conference on Green Computing and Internet of Things. 2015 Presented at: ICGCIoT'15; October 8-10, 2015; Noida, India. [doi: 10.1109/icgciot.2015.7380509]

45. Ashique K, Kaliyadan F, Aurangabadkar SJ. Clinical photography in dermatology using smartphones: an overview. Indian Dermatol Online J 2015;6(3):158-163 [FREE Full text] [doi: 10.4103/2229-5178.156381] [Medline: 26009708]

46. Połap D, Winnicka A, Serwata K, Kęsik K, Woźniak M. An intelligent system for monitoring skin diseases. Sensors (Basel) 2018 Aug 4;18(8):- [FREE Full text] [doi: 10.3390/s18082552] [Medline: 30081540]

47. Shimizu E, Ogawa Y, Yazu H, Aketa N, Yang F, Yamane M, et al. 'Smart Eye Camera': an innovative technique to evaluate tear film breakup time in a murine dry eye disease model. PLoS One 2019;14(5):e0215130 [FREE Full text] [doi: 10.1371/journal.pone.0215130] [Medline: 31071120]

48. Li S, Xu LD, Zhao S. 5G internet of things: a survey. J Ind Inf Integration 2018 Jun;10:1-9. [doi: 10.1016/j.jii.2018.01.005]

49. Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. J Ambient Intell Human Comput 2017 Dec 29;10(10):4151-4166. [doi: 10.1007/s12652-017-0659-1]

50. Pan J, McElhannon J. Future Edge Cloud and Edge Computing for Internet of Things Applications. IEEE Internet Things J 2018 Feb;5(1):439-449. [doi: 10.1109/jiot.2017.2767608]

51. Chen HS, Jarrell JT, Carpenter KA, Cohen DS, Huang X. Blockchain in healthcare: a patient-centered model. Biomed J Sci Tech Res 2019 Aug 8;20(3):15017-15022. [doi: 10.26717/bjstr.2019.20.003448]

52. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]

53. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]

54. Artificial Intelligence in Medicine. IBM. 2018. URL: https://www.ibm.com/watson-health/learn/artificial-intelligence-medicine [accessed 2020-02-02]

55. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J Am Med Assoc 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]

56. Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep learning intervention for health care challenges: some biomedical domain considerations. JMIR Mhealth Uhealth 2019 Aug 2;7(8):e11966 [FREE Full text] [doi: 10.2196/11966] [Medline: 31376272]

57. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. J Biomed Inform 2017 May;69:218-229 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.001] [Medline: 28410981]

58. Cheng Y, Wang F, Zhang P, Hu J. Risk Prediction With Electronic Health Records: A Deep Learning Approach. In: Proceedings of the 2016 SIAM International Conference on Data Mining. 2016 Presented at: SIAM'16; May 5-7, 2016; Florida, USA. [doi: 10.1137/1.9781611974348.49]

59. Yadav S, Ekbal A, Saha S, Bhattacharyyaditors. Deep Learning Architecture for Patient Data De-identification in Clinical Records. In: Proceedings of the Clinical Natural Language Processing Workshop. 2016 Presented at: ClinicalNLP'16; December 11, 2016; Osaka, Japan. [doi: 10.4018/978-1-7998-0951-7.ch048]

60. Li H, Li X, Ramanathan M, Zhang A. Identifying informative risk factors and predicting bone disease progression via deep belief networks. Methods 2014 Oct 1;69(3):257-265. [doi: 10.1016/j.ymeth.2014.06.011] [Medline: 24979059]

61. Suk H, Lee S, Shen D, Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 2014 Nov 1;101:569-582 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.06.077] [Medline: 25042445]

XSL•FO

RenderX

62. Lu N, Li T, Ren X, Miao H. A deep learning scheme for motor imagery classification based on restricted boltzmann machines. IEEE Trans Neural Syst Rehabil Eng 2017 Jun;25(6):566-576. [doi: 10.1109/TNSRE.2016.2601240] [Medline: 27542114]

63. Lee C, Chen G, Zhang Z, Chou Y, Hsu C. Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network? J Healthc Eng 2018;2018:8413403 [FREE Full text] [doi: 10.1155/2018/8413403] [Medline: 30651947]

64. Mathews SM, Kambhamettu C, Barner KE. A novel application of deep learning for single-lead ECG classification. Comput Biol Med 2018 Aug 1;99:53-62. [doi: 10.1016/j.compbiomed.2018.05.013] [Medline: 29886261]

65. Tian S, Yang W, Grange JM, Wang P, Huang W, Ye Z. Smart healthcare: making medical care more intelligent. Global Health J 2019 Sep;3(3):62-65. [doi: 10.1016/j.glohj.2019.07.001]

66. Yu L, Lu Y, Zhu X. Smart Hospital based on internet of things. J Netw 2012 Oct 1;7(10):1654-1661. [doi: 10.4304/jnw.7.10.1654-1661]

67. Thangaraj M, Ponmalar P, Anuradha S. Internet of Things (IoT) Enabled Smart Autonomous Hospital Management System - a Real World Health Care Use Case With the Technology Drivers. In: IEEE International Conference on Computational Intelligence and Computing Research. 2015 Presented at: IEEE'15; December 10-12, 2015; Madurai, India. [doi: 10.1109/iccic.2015.7435678]

68. Australian Burden of Disease Study: Impact and Causes of Illness and Death in Australia 2015. Australian Institute of Health and Welfare. 2015. URL: https://www.aihw.gov.au/reports-data/health-conditions-disability-deaths/burden-of-disease/overview [accessed 2020-10-02]

69. de Silva D. Evidence: Helping People Help Themselves. The Heart Foundation. 2011. URL: https://www.health.org.uk/publications/evidence-helping-people-help-themselves [accessed 2020-02-02]

70. Vickery DM, Kalmer H, Lowry D, Constantine M, Wright E, Loren W. Effect of a self-care education program on medical visits. J Am Med Assoc 1983 Dec 2;250(21):2952-2956. [Medline: 6358552]

71. Pai A. Survey: Percent of Smartphone Users Have Downloaded a Fitness or Health App. Mobile Health News. 2015. URL: https://www.mobihealthnews.com/48273/survey-58-percent-of-smartphone-users-have-downloaded-a-fitness-or-health-app [accessed 2020-04-04]

72. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]

73. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]

74. Bates DW, Landman A, Levine DM. Health apps and health policy: what is needed? J Am Med Assoc 2018 Nov 20;320(19):1975-1976. [doi: 10.1001/jama.2018.14378] [Medline: 30326025]

75. Borycki E. Quality and safety in ehealth: the need to build the evidence base. J Med Internet Res 2019 Dec 19;21(12):e16689 [FREE Full text] [doi: 10.2196/16689] [Medline: 31855183]

76. Pacheco R, Dias, Santinha, Rodrigues, Queirós, Rodrigues. Smart cities and healthcare: a systematic review. Technologies 2019 Aug 16;7(3):58. [doi: 10.3390/technologies7030058]

77. Lai Y, Yeung W, Celi LA. Urban intelligence for pandemic response: viewpoint. JMIR Public Health Surveill 2020 Apr 14;6(2):e18873 [FREE Full text] [doi: 10.2196/18873] [Medline: 32248145]

78. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing covid-19 among 627,386 persons in contact with the diamond princess cruise ship passengers who disembarked in Taiwan: big data analytics. J Med Internet Res 2020 May 5;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]

79. Wray A, Olstad DL, Minaker LM. Smart prevention: a new approach to primary and secondary cancer prevention in smart and connected communities. Cities 2018 Sep;79:53-69. [doi: 10.1016/j.cities.2018.02.022]

80. Palmieri F, Ficco M, Pardi S, Castiglione A. A cloud-based architecture for emergency management and first responders localization in smart city environments. Comput Electric Eng 2016 Nov;56:810-830. [doi: 10.1016/j.compeleceng.2016.02.012]

81. Korzun D. Internet of things meets mobile health systems in smart spaces. In: Internet of Things and Big Data Technologies for Next Generation Healthcare. Switzerland: Springer; 2017.

82. Dauwed M, Meri A. In: Ismail Y, editor. IOT Service Utilisation in Healthcare. New York, USA: IntechOpen; 2019.

83. World Health Organization. Global Diffusion of Ehealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on Ehealth. Slitzerland: World Health Organization; 2017.

84. Riazul Islam SM, Daehan K, Humaun Kabir M, Hossain M, Kyung-Sup Kwak. The internet of things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/access.2015.2437951]

85. Global Internet of Things Policies 2018-2019. Research and Markets. 2019. URL: https://www.globenewswire.com/news-release/2019/05/30/1859482/0/en/Global-Internet-of-Things-Policies-2018-2019.html [accessed 2020-06-04]

86. Horizon Scanning Series: An overview. Australian Council of Learned Academies. 2018. URL: https://acola.org/programs/horizon-scanning-series/ [accessed 2020-04-02]

87. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176 [FREE Full text] [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]

88. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]

89. Schofield P, Shaw T, Pascoe M. Toward comprehensive patient-centric care by integrating digital health technology with direct clinical contact in Australia. J Med Internet Res 2019 Jun 4;21(6):e12382 [FREE Full text] [doi: 10.2196/12382] [Medline: 31165713]

90. Jalali MS, Kaiser JP, Siegel M, Madnick S. The internet of things promises new benefits and risks: a systematic analysis of adoption dynamics of IoT products. IEEE Secur Privacy 2019 Mar;17(2):39-48. [doi: 10.1109/msec.2018.2888780]

91. Lee H, Park YR, Kim H, Kang NY, Oh G, Jang I, et al. Discrepancies in demand of internet of things services among older people and people with disabilities, their caregivers, and health care providers: face-to-face survey study. J Med Internet Res 2020 Apr 15;22(4):e16614 [FREE Full text] [doi: 10.2196/16614] [Medline: 32293575]

92. Verloo H, Kampel T, Vidal N, Pereira F. Perceptions about technologies that help community-dwelling older adults remain at home: qualitative study. J Med Internet Res 2020 Jun 4;22(6):e17930 [FREE Full text] [doi: 10.2196/17930] [Medline: 32496197]

93. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

94. NIST Releases Draft Security Feature Recommendations for IoT Devices. National Institute of Standards and Technology. 2019. URL: https://www.nist.gov/news-events/news/2019/08/nist-releases-draft-security-feature-recommendations-iot-devices [accessed 2020-02-02]

95. Rubí JN, Gondim PR. Iomt platform for pervasive healthcare data aggregation, processing, and sharing based on onem2m and openehr. Sensors (Basel) 2019 Oct 3;19(19):- [FREE Full text] [doi: 10.3390/s19194283] [Medline: 31623304]

96. Rubí JN, Gondim PR. Interoperable internet of medical things platform for e-health applications. Int J Distrib Sens N 2020 Jan 7;16(1):155014771988959. [doi: 10.1177/1550147719889591]

97. Desai P, Sheth A. Semantic Gateway as a Service Architecture for IoT Interoperability. In: IEEE International Conference on Mobile Services. 2015 Presented at: CMS'15; June 27-July 2, 2015; New York, USA. [doi: 10.1109/mobserv.2015.51]

98. Milosevic Z, Bond A. Digital Health Interoperability Frameworks: Use of RM-ODP Standards. In: 20th International Enterprise Distributed Object Computing Workshop. 2016 Presented at: EDOCW'16; September 5-9, 2016; Vienna, Austria. [doi: 10.1109/edocw.2016.7584359]

99. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]

Abbreviations

AI: artificial intelligence
EMR: electronic medical record
IoT: Internet of Things
mHealth: mobile health
RFID: radio frequency identification

Viewpoint

Social Media Surveillance in Schools: Rethinking Public Health Interventions in the Digital Age

Colin Burke[1], MS; Cinnamon Bloss[1], PhD

University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Cinnamon Bloss, PhD
University of California, San Diego
9500 Gilman Drive MC0896
La Jolla, CA, 92093-0896
United States
Phone: 1 8585349595
Email: cbloss@ucsd.edu

Abstract

Growing public concern about student safety and well-being has led schools and school districts to contract private companies to implement new technologies that target and surveil students' activity on social media websites. Although innovative solutions for addressing student safety and health are needed, it is unclear whether the implementation of social media surveillance in schools is an effective strategy. Currently, there is no evidence to support the claims made by social media surveillance companies, as well as the schools that hire them, that these technologies can address the myriad of public health issues facing today's students. Instead, these digital surveillance systems may only serve to exacerbate the problems that youth—especially those from historically marginalized groups—already face.

Over the last two decades, schools have faced several concerning trends related to student safety and well-being, including acts of violence, cyberbullying, and adolescent suicidality. The number of shooting incidents in K-12 schools, for instance, has increased since 1970, with a record 97 shooting incidents in 2018 alone [1]. Furthermore, emerging data indicate worrisome trends related to online bullying [2] and adverse mental health outcomes, including increased rates of adolescent suicide [3]. Since social media websites have become a primary medium for students to express their thoughts, views, and feelings, social media has also been increasingly seen as a potential site for intervention and prevention of these public health threats.

Concern about the safety and security of schools and students has led to an increasing number of US schools and school districts hiring private companies to monitor students' social media activity [4]. Companies that provide social media surveillance services purportedly have the ability to identify and report any public social media posts made by students that fall under predetermined categories of concern. The first and most widely covered case of social media surveillance took place in the Glendale School District in California, where the suicide of a student in 2013 prompted the district to contract an external company to monitor and analyze students' social media accounts. Since then, thousands of schools and school districts have hired companies to provide social media surveillance services. Although innovative solutions for addressing students' safety and health are needed, it is unclear whether the implementation of social media surveillance in schools is an effective strategy. Despite the increased implementation of social media surveillance in schools, the public is generally unaware that these services are so widely utilized. This lack of awareness means that there has also been little to no consideration of the consequences of implementing these social media surveillance technologies. More thoughtful debate and study are needed to bring about greater public and scholarly attention to the use of these technologies and to better understand the potential implications of their use in schools.

Schools and school districts are not alone in their attempts to monitor students' social media activity. Most recently, the state of Florida contracted a private technology firm called FivePoint Solutions to monitor and analyze the Florida Schools Safety Portal (FSSP), which consolidates data from Florida's Department of Education, Department of Children and Families, Department of Law Enforcement, Department of Juvenile

Justice, and local law enforcement, as well as students' posts from social media websites. The FSSP, which is currently in use by all Florida public schools, aims to allow "school threat assessment teams to identify, assess and provide intervention services for individuals whose behavior may pose a threat to themselves or others" and "report suspicious activities to the proper authorities within the school district" [5].

Taken at face value, social media surveillance services may present an opportunity for schools to increase their awareness of students' online activity, as well as better identify and prevent potential instances of harm that may otherwise go unreported. The integration of multiple databases (eg, the FSSP [Florida Schools Safety Portal]) provides school officials with a large amount of information about their students. Theoretically, this helps school officials to make more informed decisions about potential threats to the safety of their students. A recent study by the University of Chicago Crime Lab found that social media surveillance carried out by Chicago public school officials, in conjunction with targeted interventions, led to positive outcomes for Chicago public school students, including lower risk of exposure to out-of-school shooting incidents, fewer misconduct incidents, fewer out-of-school suspensions, and higher rates of school attendance [6]. Although this report suggests that social media surveillance may have some utility, there is currently no evidence that social media surveillance is able to effectively address the public health issues that many social media surveillance services, as well as the schools that use them, claim to be targeting, such as cyberbullying, students' mental health, and violence in schools.

The deployment of social media surveillance technologies in school settings also raises some concerns. First, while students' social media posts are public, educational professionals are certainly not in students' imagined audience. Considerable harm can be done through the act of sharing students' online activity with school administrators who hold positions of power over students. School administrators are not immune to bias, and simply seeing this information can potentially influence their perceptions of students in a negative way, whether consciously or subconsciously. Students may also experience embarrassment or shame as a result of knowing that their online activity was accessed by and distributed to an unintended audience, potentially exacerbating any existing mental health issues (eg, depression and anxiety). In addition, the language and culture of today's students (ie, memes) may not be easily understood and interpreted by older school officials, who are faced with the difficult task of translating and interpretating students' online activity to identify potential threats and harm. Thus, there is a considerable risk of false positives due to this cross-generational cultural barrier, which may lead to the unwarranted punishment of students, as well as general distrust in school administrators and the institutions they represent.

Second, social media surveillance technologies are vulnerable to algorithmic biases that may disproportionately target particular individuals or groups. The FSSP, for example, collects information from several Florida state-wide databases, and biases within these databases must be taken into account. The dynamics of inclusion and exclusion within these databases can have a significant impact on who is labeled a threat and subjected to increased monitoring and surveillance. This becomes especially problematic when some databases are inherently overrepresented by marginalized racial and socioeconomic groups. Black youth, for example, represent over half of the juvenile arrests in the Florida Department of Juvenile Justice database [7], which is where crime and other adolescent misconduct are logged. The overrepresentation of Black youth in this database could lead an algorithm to produce biased findings, such as considering Black youth a greater threat than members of other racial groups. Similarly, data from the Florida Department of Children and Families are more likely to include students from lower socioeconomic backgrounds. The use of these incomplete databases means that threat assessment and the resulting target of surveillance systems may be primarily directed toward students in marginalized groups who are overrepresented in such databases. In contrast, students from more privileged racial and socioeconomic groups, who these databases tend to exclude, may be less likely to be subjected to further surveillance. Algorithms are only as good as the data they are trained on; feeding algorithm-biased data leads to biased outcomes [8-10], which, in this case, could mean disproportionately targeting Black youth and students from lower socioeconomic backgrounds.

The Florida case highlights the possible dangers that are inherent in relying upon digital technological systems and so-called "big data" to inform public health interventions without careful consideration of the limitations and biases that exist within these systems. Although innovative approaches to student safety and well-being are needed, it is unclear whether the implementation of social media surveillance in schools is an effective strategy. Despite their potential utility, there is currently no compelling empirical evidence to support the claims made by social media surveillance companies, as well as the schools that hire them, that social media surveillance technologies can effectively address the public health threats facing today's students, such as cyberbullying, adolescent suicidality, and acts of violence. Instead, these digital surveillance systems may only serve to exacerbate the problems that youth—particularly those from historically marginalized groups—already face. Furthermore, these public health interventions are being carried out on a massive scale with little public awareness and regard for the consequences of their implementation. The potential pitfalls of social media surveillance mean that we, as a society, must engage in a more extensive dialogue to focus our attention on the use and effects of these technologies.

Conflicts of Interest

None declared.

References

1. K-12 School Shooting Database. Incidents by Year. Shooting Incidents at K-12 Schools 1970-Present.: Center for Homeland Defense and Security URL: https://www.chds.us/ssdb/charts-graphs/#incedentsbyyear [accessed 2020-10-18]
2. Anderson M, Smith A, Nolan H. A Majority of Teens Have Experienced Some Form of Cyberbullying. Pew Research Center.: Pew Research Center; 2018 Sep. URL: https://www.pewinternet.org/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/ [accessed 2020-10-18]
3. Curtin SC, Warner M, Hedegaard H. Increase in Suicide in the United States, 1999-2014. National Center for Health Statistics. Hyattsville, Maryland: National Center for Health Statistics; 2016 Apr. URL: https://www.cdc.gov/nchs/products/databriefs/db241.htm [accessed 2020-10-18]
4. Shade LR, Singh R. "Honestly, We're Not Spying on Kids": School Surveillance of Young People's Social Media. Social Media + Society 2016 Nov 29;2(4):1-12. [doi: 10.1177/2056305116680005]
5. Reid S, Braun J. Florida Department of Education.: Florida Department of Education; 2019 Aug 15. URL: http://www.fdle.state.fl.us/MSDHS/Meetings/2019/August/August-15-200pm-Florida-Schools-Safety-Presentatio.aspx [accessed 2020-10-19]
6. University of Chicago Crime Lab. University of Chicago Crime Lab. Chicago, Illinois; 2019 Jan. URL: https://www.ncjrs.gov/pdffiles1/nij/grants/252718.pdf [accessed 2020-10-19]
7. Florida Department of Juvenile Justice. Delinquency Profile Dashboard. Juvenile Justice Information System. URL: http://www.djj.state.fl.us/research/reports/reports-and-data/interactive-data-reports/delinquency-profile/delinquency-profile-dashboard [accessed 2020-10-19]
8. Benjamin R. Race After Technology: Abolitionist Tools for the New Jim Code. Cambridge, UK: Polity Press; 2019.
9. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Press; Jan 23, 2018.
10. Noble SU. Algorithms of Oppression: How Search Engines Reinforce Racism. New York, NY: NYU Press; 2018.

Abbreviations

FSSP: Florida Schools Safety Portal

XSL•FO

RenderX

Original Paper

Mobile App Strategy to Facilitate Human Papillomavirus Vaccination Among Young Men Who Have Sex With Men: Pilot Intervention Study

Holly B Fontenot[1,2,3], PhD, RN, NP; Bradley Patrick White[3], MSN, RN; Joshua G Rosenberger[4], PhD; Hailee Lacasse[3], BSN, RN; Chokdee Rutirasiri[5,6], BSBA; Kenneth H Mayer[2,7,8], MD; Gregory Zimet[9], PhD

[1]School of Nursing and Dental Hygiene, University of Hawaii at Manoa, Honolulu, HI, United States

[2]The Fenway Institute, Boston, MA, United States

[3]Connell School of Nursing, Boston College, Chestnut Hill, MA, United States

[4]College of Health and Human Development, Penn State University, University Park, PA, United States

[5]Proper Villains, Boston, MA, United States

[6]School of Arts and Sciences, Boston College, Chestnut Hill, MA, United States

[7]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[8]School of Medicine, Harvard University, Boston, MA, United States

[9]Division of Adolescent Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States

Corresponding Author:
Holly B Fontenot, PhD, RN, NP
School of Nursing and Dental Hygiene
University of Hawaii at Manoa
2528 McCarthy Mall
Webster Hall 412
Honolulu, HI, 96822
United States
Phone: 1 808 956 4917
Email: hbfont@hawaii.edu

Abstract

Background: Mobile app-based interventions have been identified as potential facilitators for vaccination among young men who have sex with men (MSM).

Objective: This pilot study aimed to test the feasibility of a theoretically informed mobile health (mHealth) tool designed to reduce health disparities and facilitate human papillomavirus (HPV) vaccination among a sample of young MSM.

Methods: The development of the mHealth tool was guided by previous research, implementation intention theory, and design thinking. We recruited MSM aged 18-26 years through a popular online dating app and linked participants to our mHealth tool, which provided HPV vaccine information and fostered access to care.

Results: A total of 42 young MSM participated in this pilot study in Boston, Massachusetts. Participants reported variable HPV knowledge (ie, high knowledge of HPV risk factors and low knowledge of HPV-related cancer risks for men) and positive vaccine beliefs and attitudes. Of those who were either unvaccinated, not up to date, or did not report vaccine status, 23% (8/35) utilized the mHealth tool to obtain HPV vaccination. Participants primarily utilized the tool's (1) educational components and (2) capabilities facilitating concrete vaccine action plans.

Conclusions: We recruited an underserved at-risk population of youth via an online dating app for our mHealth intervention that resulted in in-person health care delivery. This study was limited by enrollment challenges, including low willingness to download the mHealth tool to mobile devices.

KEYWORDS

XSL·FO

RenderX

Introduction

In the United States, human papillomavirus (HPV) infection is a significant public health issue, and a disparity exists for men who have sex with men (MSM). As compared to men, in general, MSM experience a higher prevalence of HPV and HPV-related diseases [1]. HPV-related diseases among men include anogenital warts, anal cancer, penile cancer, oropharyngeal warts, and oropharyngeal cancer [2,3]. US national rates of HPV-related cancers are increasing among men [4,5], and rates of HPV-related anal cancers are the highest among young MSM [2,5,6].

Despite the availability of HPV vaccination for men since 2009, rates of HPV vaccination initiation and series completion among young MSM remain unacceptably low in the United States. In 2018, when this study was conducted in the United States, national vaccination rates (≥ 1 dose) in 13- to 17-year-old adolescent boys were low (41.7%) [7], and the rates were notably lower among young MSM (up to age 26 years; reported rates ranged from 4.9% to 13%) [8-10]. These rates contrasted sharply with the national rates of HPV vaccination in women at the time (60% adolescent girls aged 13-17 years had received ≥1 HPV vaccine dose) [7].

There are several factors that contribute to HPV vaccine disparities for men and young MSM, in particular. It is likely that the initial marketing of the cervical cancer vaccine for women combined with weak recommendations (or no recommendations) from health care providers for HPV vaccination for men led to these disparities among all men [11]. The US national HPV guidelines (from 2009 to 2019) recommended vaccination in a 3-dose series for all men up to age 21 years and up to age 26 years for young MSM. However, this sexual orientation–based vaccine recommendation was problematic, as (1) many young MSM (due to stigma and discrimination) do not disclose their sexual orientation to their health care providers [12,13] and (2) health providers do not routinely ask about sexual orientation [14]. This lack of open communication (assessment and disclosure) decreases the likelihood that appropriate health services are offered [12,15]. Lastly, young MSM also report low knowledge on HPV and the HPV vaccine, a lack of awareness of vaccine indications for men, and low perceived risk of HPV acquisition [16-18].

Overall, research on barriers to vaccination indicates that both attitudinal (eg, health beliefs) and logistical (eg, health care provider recommendation and concerns about cost) factors interfere with vaccine uptake [8,19,20]. Previous research indicates that young MSM report positive vaccine attitudes along with a belief that mobile app-based interventions can facilitate vaccination and access to care [16,21,22].

Young MSM tend to heavily utilize social media and mobile apps to seek sexual health information as well as sexual partners [23-25]. There is a rapid growth of mobile apps designed to facilitate meeting social/sexual partners, with recent estimates indicating at least 6.2 million MSM users nationally [23]. These extremely popular mobile social networks represent "virtual" communities of young MSM. Mobile social network/dating apps have been used to successfully recruit MSM for rectal microbicide [22] and sexual risk behavior research [23]. In addition, young MSM have overwhelmingly identified a willingness to participate in research if advertised on MSM mobile apps [18,21]. Therefore, recruitment of young MSM via a mobile social network for an app-based intervention [24] is an important strategy to consider for HPV vaccine promotion.

This pilot study evaluated a novel app-focused approach to facilitate HPV vaccination among young MSM. We connected with young MSM through popular dating apps and linked them to our mobile health (mHealth) tool, a web-enabled app designed to reduce MSM-specific barriers to vaccination (eg, knowledge/awareness, stigma, and discrimination) and overcome logistical barriers to HPV vaccination (eg, access to MSM-affirming health care and cost concerns).

Methods

mHealth Tool and Theory

The mHealth tool is a universally compatible web-enabled mobile app that was developed for this pilot study based on our previous research with the target audience of young MSM [18], implementation intention theory [26], and design thinking. The implementation intention theory proposes that the gap between intention and behavior can be effectively bridged by concrete action plans that address when, where, and how the intention can be translated into action [26-28]. Design thinking is a human-centered framework that places the target audience (end user of our mHealth tool) at the center of all key design decisions. Design thinking utilizes empathy to guide a collaborative, codesign process centered on end user needs, wants, and behaviors to create solutions that engage the target audience [29]. The mHealth tool sought to bridge the intention-behavior gap by addressing known barriers to vaccination among young MSM [12,16-18,30]. The tool also served as a bridge between the virtual community and an MSM-affirming health center. Informed by young MSM [18], the tool content included education on HPV, HPV vaccine, prevention of HIV/sexually transmitted infection, and how the health center could help with health insurance enrollment and vaccine cost assistance. Tool functionalities included geolocation/directions to the health center, appointment scheduling, and appointment reminder system.

Sample and Setting

A convenience sample of young MSM aged 18-26 years was recruited over 6 months (March 2018 to September 2018) via banner and push advertisements primarily posted on 1 mobile MSM dating app that is popular among racially/ethnically diverse men. All advertisements were geolocated to app users in Boston, Massachusetts. Toward the end of the recruitment timeframe, we trialed an additional 1 month of advertisements with another well-known MSM dating app to boost enrollment efforts (our goal was to enroll 200 young MSM). Young MSM who clicked on the study advertisement were directed to the study webpage to assess eligibility and provide informed consent. Participants were eligible if they identified as MSM, were aged 18-26 years, and were a user of our targeted MSM dating app in the Boston area. Eligible and consented individuals were invited to download the study mHealth tool (iOS and

Android compatible) to their mobile devices. At the time of download, participants were invited to complete an electronic study questionnaire that collected demographic data and measured knowledge, attitudes, and beliefs related to HPV vaccination. Participants either completed this questionnaire during the initial tool download or returned to completing the questionnaire at the time of their choosing after download (this option was made available to reduce barriers for those for whom time was a constraint).

The partner health center was prepared to receive study participants, enroll them into care, assist with insurance enrollment or vaccine cost assistance programs, and provide HPV vaccine services as medically indicated. HPV vaccination appointments made in the mHealth tool (monitored by the principal investigator) were communicated directly to the health center administrator so that an appointment could be booked as "study participant–HPV vaccination" (as personal identifiers were not collected in the mHealth tool). Once an individual arrived at the center for the scheduled appointment, they identified themselves, and usual care commenced. This health center primarily cares for sexual and gender minority adolescents and adults up to age 29 years, and appointments are largely booked as "same-day" or "walk-in." Nurses (or other health providers) provided usual vaccine care and education as per the clinic protocol. Participants receiving HPV vaccination at the health center were asked to log their clinic visit on the study-specific clinic iPad. No personal identifiers were collected. Study protocols and procedures were approved by the Fenway Health Institutional Review Board.

Measures

We sequentially collected the following rates of participation in the study.

1. Accessing the study webpage (collected via Qualtrics)
2. mHealth tool download and usage data (collected within the study tool)
3. Initiation/completion of study questionnaire (collected via Qualtrics)
4. HPV vaccine initiation (collected via study-specific clinic iPad that was linked to Qualtrics or via self-report in the mHealth tool by those who obtained vaccination with another health provider)

The study questionnaire included questions on demographics (eg, race/ethnicity, health insurance status, and employment/student status) as well as previously validated questions on HPV knowledge, attitudes toward HPV and HPV vaccination, subjective norms, perceived behavioral controls,

and degree of intention toward vaccination, adapted from our prior work [31-34]. Participants were offered a US $5 gift card plus entry into a prize drawing for a US $75 gift card (1 awarded each month) for completing the questionnaire.

Data Analysis

The primary outcomes were rates of participation in each sequential study step. Data collected via Qualtrics and through the mHealth tool were managed using Microsoft Excel. Descriptive analyses of study steps (absolute and intermediate completion rates) were completed. Other descriptive analyses included means for continuous variables and percentages for nominal data.

Results

At the time of recruitment initiation (March 2018), available data from the partnered dating app indicated that approximately 2021 Massachusetts-based users aged 18 to 25 years had logged in during the previous 30 days. During the 6 months of recruitment, a total of 338 potential participants accessed the study web page and engaged in some way. Out of those potential participants, 54 participants met the eligibility requirements and provided informed consent. A total of 42 participants then chose to download the mHealth tool on their mobile devices and participate in the pilot study. Out of these, 33 completed the study questionnaire (Figure 1) and 67% (22/33) completed the questionnaire during the initial mHealth tool download.

The mean age of participants who completed the study questionnaire was 22.7 years. The sample was diverse. For example, 18% (6/33) of the participants reported ethnicity as Hispanic. Race was reported as 9% (3/33) Asian, 6% (2/33) American Indian/Alaska Native, 18% (6/33) Black, 18% (6/33) White, and 18% (6/33) Multiracial; 30% (10/33) participants did not report their race. Out of the 33 participants, 4 (12%) completed some high school, 5 (15%) were high school graduates, 6 (18%) attended some college, 6 (18%) were college graduates, and 36% (12/33) did not report their education level. Health insurance type was reported as 24% (8/33) private, 24% (8/33) public, and 12% (4/33) uninsured; 40% (13/33) participants did not report on their health insurance. Lastly, 30% (10/33) worked full-time; 18% (6/33) were full-time students; 9% (3/33) were part-time students or worked part-time; and 42% (14/33) did not report on their work/student status.

Participants reported variable HPV knowledge (ie, higher knowledge of HPV risk factors and lower knowledge of HPV-related cancer risks for men) and generally positive vaccine beliefs and attitudes (Table 1).

Figure 1. Participant progression though study protocol. HPV: human papillomavirus; mobile health: mHealth.
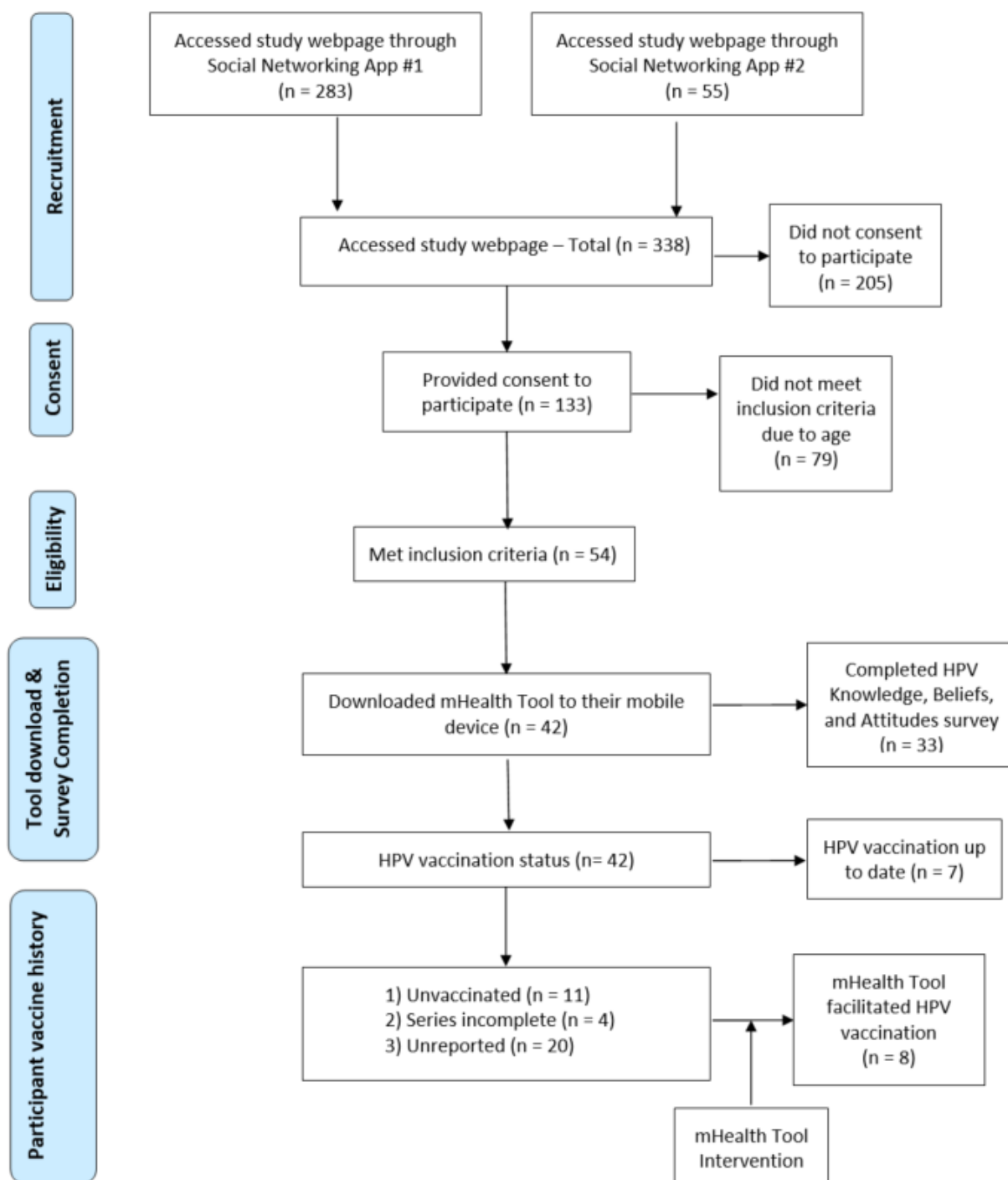
Of the 42 participants who engaged with the mHealth Tool, 7 (17%) were up-to-date with their HPV vaccination, 11 (26%) were unvaccinated, 4 (9%) were not up-to-date on vaccine dosing, and 20 (48%) did not report their vaccination status. Of those who were either unvaccinated, not up-to-date, or did not report status, 23% (8/35) utilized the mHealth tool to facilitate HPV vaccine. Out of these 8 participants, 4 obtained vaccine at the partnered health center, and 4 obtained vaccine at another health center (Figure 1).

Usage data from the mHealth tool revealed that participants primarily accessed (1) the tool's educational components focused on HPV and the HPV vaccine and (2) the tool's functionalities that facilitated development of concrete action plans (eg, appointment booking, appointment reminders, and health center geolocation directions).

Table 1. Questionnaire data on HPV knowledge, attitudes, and beliefs.[a]

Questionnaire items	Responses, n/N (%)	Responses, n/N (%)
HPV[b] knowledge	Yes	No
Have you ever heard of the HPV?	24/32 (75)	8/32 (25)
Has your health care provider ever recommended the HPV vaccine?	19/31 (61)	12/31 (39)
Attitudes	True	False
The use of condoms can help protect you from getting HPV	19/23 (83)	4/23 (17)
HPV can be spread through oral to genital contact.	19/23 (83)	4/23 (17)
A person may have HPV and not know it.	21/22 (95)	1/22 (5)
Both men and women can get HPV.	21/23 (91)	2/23 (9)
HPV can cause anal cancer.	13/22 (59)	9/22 (41)
HPV can be cured.	12/22 (54)	10/22 (46)
HPV can cause cancers of the head and neck.	13/22 (59)	9/22 (41)
Having a history of multiple sexual partners increases your risk for HPV.	20/22 (91)	2/22 (9)
Attitudes and Beliefs	Agree	Disagree/neutral
Vaccines are important to prevent disease that can be spread person to person.	17/19 (89)	2/19 (11)
Vaccines should be required for contagious diseases that can be spread person to person.	18/18 (100)	0/18 (0)
If there was a vaccine that prevented cancer, I would want it.	15/18 (83)	3/18 (17)
It is safe to get vaccinated for HPV.	14/18 (78)	4/18 (22)
The HPV vaccine only helps women.	3/18 (17)	15/18 (83)
It is important for gay men to get the HPV vaccine.	15/18 (83)	3/18 (17)
My parents would think less of me if I got the HPV vaccine.	4/16 (25)	12/16 (75)
My friends would think less of me if I got the HPV vaccine.	2/16 (12)	14/16 (88)
My sexual partners would think less of me if I got the HPV vaccine.	2/16 (12)	14/16 (88)
It would be hard for me to find time to get vaccinated for HPV.	4/16 (25)	12/16 (75)
It would be hard for me to get the HPV vaccine because I am afraid of shots.	5/16 (31)	11/16 (69)
It would be hard for me to find a health center/health provider in order to get vaccinated.	3/16(19)	13/16 (81)
I am planning on getting the HPV vaccine within the next 30 days	6/16 (38)	10/16 (62)

[a]Findings are reported as n/N (%), where N is the number of people who answered each question, which varied from question to question.

[b]HPV: human papillomavirus.

Discussion

We aimed to determine if it was possible to recruit young MSM via a dating app for an mHealth (app-based) intervention that would effectively engage young MSM in local health services related to HPV vaccination. The results of this study support the theoretical framework (implementation intention theory), which posits that helping youth create concrete action plans can facilitate health behaviors. Future uses of mHealth tools may be effective in bridging the intention-behavior gap, particularly among adolescents (native technology users).

Of the participants who engaged with the mHealth tool and were either unvaccinated, not up-to-date on vaccine series completion, or did not report their vaccine status, 23% (8/35) utilized the tool to facilitate vaccination with the partnered health center or elsewhere. App usage data documented that participants primarily accessed information about HPV vaccination and used the tools to schedule appointments, geolocate the clinic, and set appointment reminders. Although there are few comparison mHealth studies among young MSM, another web-based educational and email/text message reminder intervention noted success in increasing HPV vaccination rates [35].

Although this study piloted a novel concept for engaging young MSM and for increasing their access to an affirming health center to facilitate HPV vaccination, the study had several limitations. We encountered challenges in study enrollment, and the study was limited by its narrow geographic focus. Enrollment was challenged by participants' willingness to download the study mHealth tool and engage in the study questionnaire. Thus, we could engage only 42 participants and not 200 as per our original goal. Despite past success engaging young MSM to participate in nonlongitudinal research using dating apps [18], it is possible that many such users may not

want to be redirected to an external website or be willing to download a study-specific web-based app where they would be asked to engage over time. Future mHealth research and interventions with young MSM may benefit from adjusted recruitment (eg, recruitment on Facebook or Instagram) and engagement strategies (eg, website-based rather than app-based), which have shown greater evidence of success [35,36] nationally.

Research highlighting varied uses of mHealth tools targeting adolescents and young adults to improve HPV vaccination as well as reduce sexually transmitted infections shows promise [35,37,38]. Our findings suggest that an mHealth tool to support young MSM to create concrete action plans and bridge the intention-behavior gap is possible and may be useful when the health behavior has associated stigmas and when the behavioral goal is singular (eg, HPV vaccination). However, future research is needed to optimize the recruitment and longitudinal engagement strategy of an mHealth tool so that the desired outcomes (eg, increasing HPV vaccination) can be optimally met. Future research should also focus on states or regions with fewer health resources and lower baseline vaccine rates than those in Massachusetts.

The study questionnaire supported previous research [16] that young MSM view HPV vaccination as an important preventative health intervention. However, similar to other studies, results indicate a continued need to enhance knowledge regarding HPV vaccine safety, appropriateness of vaccination for MSM, and HPV-associated cancers in men (eg, anal cancer) [16-18,30]. Interventions aimed at reducing stigmas and increasing access to MSM-affirming health environments that also provide inclusive health education show promise [12,16,18] and are warranted, so that young MSM can comfortably seek health care and gain knowledge of risks and complications associated with HPV disease among MSM.

This pilot study tested feasibility of a theoretically informed mHealth tool designed to reduce health disparities and facilitate HPV vaccination among a sample of young MSM in Boston. We were able to recruit an underserved at-risk population of youth via a mobile online dating platform for an app-based intervention that resulted in in-person health delivery (virtual world to the real world). Considering some of the recruitment challenges, future work is needed to identify optimal strategies to recruit and retain users of mHealth tool technologies to improve the desired outcomes.

Conflicts of Interest

GZ has served as a paid consultant to Sanofi Pasteur for work on the Adolescent Immunization Initiative. He has served as a paid consultant to Merck related to HPV vaccine hesitancy and received travel support to present at an HPV vaccine symposium. He has received research funding through his institution from Merck via their investigator-initiated studies program.

References

1. Lieblong BJ, Montgomery BEE, Su LJ, Nakagawa M. Natural history of human papillomavirus and vaccinations in men: A literature review. Health Sci Rep 2019 May;2(5):e118 [FREE Full text] [doi: 10.1002/hsr2.118] [Medline: 31139757]

2. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2018. Atlanta: U.S. Department of Health and Human Services; 2019. URL: https://www.cdc.gov/std/stats18/default.htm [accessed 2020-10-20] [WebCite Cache ID DOI: 10.15620/cdc.79370]

3. Han JJ, Beltran TH, Song JW, Klaric J, Choi YS. Prevalence of Genital Human Papillomavirus Infection and Human Papillomavirus Vaccination Rates Among US Adult Men: National Health and Nutrition Examination Survey (NHANES) 2013-2014. JAMA Oncol 2017 Jun 01;3(6):810-816 [FREE Full text] [doi: 10.1001/jamaoncol.2016.6192] [Medline: 28114440]

4. Young D, Xiao CC, Murphy B, Moore M, Fakhry C, Day TA. Increase in head and neck cancer in younger patients due to human papillomavirus (HPV). Oral Oncol 2015 Aug;51(8):727-730. [doi: 10.1016/j.oraloncology.2015.03.015] [Medline: 26066977]

5. Palefsky J. Anal HPV infection and HPV-associated disease. In: Jenkins D, Bosch X, editors. Human Papillomavirus: Proving and Using a Viral Cause for Cancer. Academic Press: London; 2020:234-446.

6. Machalek DA, Grulich AE, Jin F, Templeton DJ, Poynten IM. The epidemiology and natural history of anal human papillomavirus infection in men who have sex with men. Sex Health 2012 Dec;9(6):527-537. [doi: 10.1071/SH12043] [Medline: 23380235]

7. Reagan-Steiner S, Yankey D, Jeyarajah J, Elam-Evans LD, Singleton JA, Curtis CR, et al. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13-17 Years--United States, 2014. MMWR Morb Mortal Wkly Rep 2015 Jul 31;64(29):784-792 [FREE Full text] [doi: 10.15585/mmwr.mm6429a3] [Medline: 26225476]

8. Cummings T, Kasting ML, Rosenberger JG, Rosenthal SL, Zimet GD, Stupiansky NW. Catching Up or Missing Out? Human Papillomavirus Vaccine Acceptability Among 18- to 26-Year-old Men Who Have Sex With Men in a US National Sample. Sex Transm Dis 2015 Nov;42(11):601-606. [doi: 10.1097/OLQ.0000000000000358] [Medline: 26462183]

9. Reiter PL, McRee A, Katz ML, Paskett ED. Human Papillomavirus Vaccination Among Young Adult Gay and Bisexual Men in the United States. Am J Public Health 2015 Jan;105(1):96-102. [doi: 10.2105/AJPH.2014.302095] [Medline: 25393178]

10. Meites E, Markowitz LE, Paz-Bailey G, Oster AM, NHBS Study Group. HPV vaccine coverage among men who have sex with men - National HIV Behavioral Surveillance System, United States, 2011. Vaccine 2014 Nov 12;32(48):6356-6359 [FREE Full text] [doi: 10.1016/j.vaccine.2014.09.033] [Medline: 25258097]

11. Kardas-Nelson M. Vaccine uptake and prevalence of HPV related cancers in US men. BMJ 2019 Mar 18;364:l1210. [doi: 10.1136/bmj.l1210] [Medline: 30885904]

12. Apaydin KZ, Fontenot HB, Shtasel D, Dale SK, Borba CP, Lathan CS, et al. Facilitators of and barriers to HPV vaccination among sexual and gender minority patients at a Boston community health center. Vaccine 2018 Jun 18;36(26):3868-3875 [FREE Full text] [doi: 10.1016/j.vaccine.2018.02.043] [Medline: 29778516]

13. Meites E, Krishna NK, Markowitz LE, Oster AM. Health care use and opportunities for human papillomavirus vaccination among young men who have sex with men. Sex Transm Dis 2013 Feb;40(2):154-157 [FREE Full text] [doi: 10.1097/OLQ.0b013e31827b9e89] [Medline: 23321994]

14. Wheldon CW, Sutton SK, Fontenot HB, Quinn GP, Giuliano AR, Vadaparampil ST. Physician Communication Practices as a Barrier to Risk-Based HPV Vaccine Uptake Among Men Who Have Sex with Men. J Cancer Educ 2018 Oct;33(5):1126-1131. [doi: 10.1007/s13187-017-1223-6] [Medline: 28456947]

15. Stupiansky NW, Liau A, Rosenberger J, Rosenthal SL, Tu W, Xiao S, et al. Young Men's Disclosure of Same Sex Behaviors to Healthcare Providers and the Impact on Health: Results from a US National Sample of Young Men Who Have Sex with Men. AIDS Patient Care STDS 2017 Aug;31(8):342-347. [doi: 10.1089/apc.2017.0011] [Medline: 28753396]

16. Fontenot HB, Fantasia HC, Vetters R, Zimet GD. Increasing HPV vaccination and eliminating barriers: Recommendations from young men who have sex with men. Vaccine 2016 Dec 07;34(50):6209-6216. [doi: 10.1016/j.vaccine.2016.10.075] [Medline: 27838067]

17. Fontenot HB, Fantasia HC, Charyk A, Sutherland MA. Human papillomavirus (HPV) risk factors, vaccination patterns, and vaccine perceptions among a sample of male college students. J Am Coll Health 2014;62(3):186-192. [doi: 10.1080/07448481.2013.872649] [Medline: 24328855]

18. Fontenot HB, Rosenberger JG, McNair KT, Mayer KH, Zimet G. Perspectives and preferences for a mobile health tool designed to facilitate HPV vaccination among young men who have sex with men. Hum Vaccin Immunother 2019;15(7-8):1815-1823 [FREE Full text] [doi: 10.1080/21645515.2019.1568156] [Medline: 30625049]

19. Rambout L, Tashkandi M, Hopkins L, Tricco AC. Self-reported barriers and facilitators to preventive human papillomavirus vaccination among adolescent girls and young women: a systematic review. Prev Med 2014 Jan;58:22-32. [doi: 10.1016/j.ypmed.2013.10.009] [Medline: 24176938]

20. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to human papillomavirus vaccination among US adolescents: a systematic review of the literature. JAMA Pediatr 2014 Jan;168(1):76-82 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.2752] [Medline: 24276343]

21. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]

22. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: 10.1007/s10461-012-0277-z] [Medline: 22851153]

23. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]

24. Grov C, Breslow AS, Newcomb ME, Rosenberger JG, Bauermeister JA. Gay and bisexual men's use of the Internet: research from the 1990s through 2013. J Sex Res 2014;51(4):390-409 [FREE Full text] [doi: 10.1080/00224499.2013.871626] [Medline: 24754360]

25. Watchirs Smith L, Guy R, Degenhardt L, Yeung A, Rissel C, Richters J, et al. Meeting Sexual Partners Through Internet Sites and Smartphone Apps in Australia: National Representative Study. J Med Internet Res 2018 Dec 18;20(12):e10683 [FREE Full text] [doi: 10.2196/10683] [Medline: 30563809]

26. Gollwitzer PM, Brandstätter V. Implementation intentions and effective goal pursuit. Journal of Personality and Social Psychology 1997;73(1):186-199. [doi: 10.1037/0022-3514.73.1.186]

27. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]

28. Gollwitzer P, Sheeran P. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology 2006;38:69-119. [doi: 10.1016/s0065-2601(06)38002-1]

XSL•FO
RenderX

29. The field guide to human-centered design. IDEO. 2015. URL: http://www.designkit.org/resources/ [accessed 2020-10-20]

30. Gerend MA, Madkins K, Crosby S, Korpak AK, Phillips G, Bass M, et al. A Qualitative Analysis of Young Sexual Minority Men's Perspectives on Human Papillomavirus Vaccination. LGBT Health 2019 Oct;6(7):350-356 [FREE Full text] [doi: 10.1089/lgbt.2019.0086] [Medline: 31556791]

31. Liau A, Stupiansky NW, Rosenthal SL, Zimet GD. Health beliefs and vaccine costs regarding human papillomavirus (HPV) vaccination among a U.S. national sample of adult women. Prev Med 2012;54(3-4):277-279. [doi: 10.1016/j.ypmed.2012.02.002] [Medline: 22342703]

32. Allen JD, Mohllajee AP, Shelton RC, Othus MK, Fontenot HB, Hanna R. Stage of adoption of the human papillomavirus vaccine among college women. Prev Med 2009 May;48(5):420-425. [doi: 10.1016/j.ypmed.2008.12.005] [Medline: 19133288]

33. Kahn J, Rosenthal S, Jin Y, Huang B, Namakydoust A, Zimet G. Rates of human papillomavirus vaccination, attitudes about vaccination, and human papillomavirus prevalence in young women. Obstet Gynecol 2008 May;111(5):1103-1110. [doi: 10.1097/AOG.0b013e31817051fa] [Medline: 18448742]

34. Fontenot HB. Intersection of HPV and sexual assault: an opportunity for practice change. J Forensic Nurs 2013;9(3):146-54; quiz E1. [doi: 10.1097/JFN.0b013e318291b276] [Medline: 24158152]

35. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing Human Papillomavirus Vaccination Among Young Gay and Bisexual Men: A Randomized Pilot Trial of the Outsmart HPV Intervention. LGBT Health 2018 Jul;5(5):325-329 [FREE Full text] [doi: 10.1089/lgbt.2018.0059] [Medline: 29979642]

36. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting Young Gay and Bisexual Men for a Human Papillomavirus Vaccination Intervention Through Social Media: The Effects of Advertisement Content. JMIR Public Health Surveill 2017 Jun 02;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]

37. Widman L, Nesi J, Kamke K, Choukas-Bradley S, Stewart J. Technology-Based Interventions to Reduce Sexually Transmitted Infections and Unintended Pregnancy Among Youth. J Adolesc Health 2018 Jun;62(6):651-660 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.02.007] [Medline: 29784112]

38. Lee HY, Koopmeiners JS, McHugh J, Raveis VH, Ahluwalia JS. mHealth Pilot Study: Text Messaging Intervention to Promote HPV Vaccination. Am J Health Behav 2016 Jan;40(1):67-76 [FREE Full text] [doi: 10.5993/AJHB.40.1.8] [Medline: 26685815]

Abbreviations

HPV: human papillomavirus
mHealth: mobile health
MSM: Men who have sex with men

Original Paper

Patients' Experiences of Telephone-Based and Web-Based Cognitive Behavioral Therapy for Irritable Bowel Syndrome: Longitudinal Qualitative Study

Stephanie Hughes[1], MSc; Alice Sibelli[2], PhD; Hazel A Everitt[1], PhD, MD; Rona Moss-Morris[2], PhD; Trudie Chalder[3], PhD; J Matthew Harvey[1], MSc; Andrea Vas Falcao[1], MSc; Sabine Landau[2], PhD; Gilly O'Reilly[1], PhD; Sula Windgassen[3], PhD; Rachel Holland[2], MSc; Paul Little[1], PhD, MD; Paul McCrone[2], PhD; Kimberley Goldsmith[2], PhD; Nicholas Coleman[4], MD; Robert Logan[5], MD; Felicity L Bishop[1], PhD

[1]Centre for Clinical and Community Applications of Health Psychology, Department of Psychology, University of Southampton, Southampton, United Kingdom

[2]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[3]Academic Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[4]Department of Gastroenterology, Southampton University Hospital, Southampton, United Kingdom

[5]Kings College Hospital, London, United Kingdom

Corresponding Author:
Felicity L Bishop, PhD
Centre for Clinical and Community Applications of Health Psychology
Department of Psychology
University of Southampton
Department of Psychology (B44), Highfield Campus
Southampton, SO17 1BJ
United Kingdom
Phone: 44 2380599020
Email: f.l.bishop@southampton.ac.uk

Abstract

Background: Cognitive behavioral therapy (CBT) is recommended in guidelines for people with refractory irritable bowel syndrome (IBS). However, the availability of CBT is limited, and poor adherence has been reported in face-to-face CBT.

Objective: Nested within a randomized controlled trial of telephone- and web-delivered CBT for refractory IBS, this qualitative study aims to identify barriers to and facilitators of engagement over time with the interventions, identify social and psychological processes of change, and provide insight into trial results.

Methods: A longitudinal qualitative study was nested in a randomized controlled trial. Repeated semistructured interviews were conducted at 3 (n=34) and 12 months (n=25) post baseline. Participants received telephone-based CBT (TCBT; n=17 at 3 months and n=13 at 12 months) or web-based CBT (WCBT; n=17 at 3 months and n=12 at 12 months). Inductive thematic analysis was used to analyze the data.

Results: Participants viewed CBT as credible for IBS, perceived their therapists as knowledgeable and supportive, and liked the flexibility of web-based and telephone-based delivery; these factors facilitated engagement. Potential barriers to engagement in both groups (mostly overcome by our participants) included initial skepticism and concerns about the biopsychosocial nature of CBT, initial concerns about telephone-delivered talking therapy, challenges of maintaining motivation and self-discipline given already busy lives, and finding nothing new in the WCBT (WCBT group only). Participants described helpful changes in their understanding of IBS, attitudes toward IBS, ability to recognize IBS patterns, and IBS-related behaviors. Consistent with the trial results, participants described lasting positive effects on their symptoms, work, and social lives. Reasons and remedies for some attenuation of effects were identified.

Conclusions: Both TCBT and WCBT for IBS were positively received and had lasting positive impacts on participants' understanding of IBS, IBS-related behaviors, symptoms, and quality of life. These forms of CBT may broaden access to CBT for IBS.

Introduction

Background

Irritable bowel syndrome (IBS) affects 10%-20% of the general population [1]. Official UK guidelines for the management of IBS [1] recommend provision of diet and lifestyle advice, a trial of medications, and—if patients have ongoing troublesome symptoms after 12 months (refractory IBS)—referral for psychological intervention, such as cognitive behavioral therapy (CBT). CBT can improve IBS symptom severity and quality of life [2-5]. However, barriers to CBT for IBS exist, including limited availability of face-to-face CBT for IBS, uncertain cost-effectiveness [6], and issues with poor adherence [7]. The Assessing Cognitive behavioral Therapy in Irritable Bowel (ACTIB) trial [8] aimed to determine the clinical effectiveness and cost-effectiveness of therapist-delivered, telephone-based CBT (TCBT) and web-based CBT (WCBT) for IBS. Both TCBT and WCBT groups showed significant improvements in IBS symptoms at 12 months, compared with treatment as usual (TAU) [9,10]. Scores from the IBS Symptom Severity Scale [11] were 61.6 (95% CI 33.8-89.5) points lower (*P*<.001) in TCBT and 35.2 (95% CI 12.6-57.8) points lower (*P*=.002) in WCBT at 12 months, than TAU.

Therapist-delivered TCBT and WCBT may overcome some of the barriers to traditional face-to-face CBT by offering better cost-effectiveness for health care commissioners [12] and providing greater flexibility in timing and location for patients [9]. Although some people may prefer remote access therapies, it may not be appropriate for all patients. Providing remote treatment options can increase access and free up more intensive face-to-face resources for those patients for whom remote intervention is not appropriate. Patients' experiences of these modalities of CBT for IBS have rarely been studied and could provide novel insights into the processes underpinning treatment uptake, adherence, and effectiveness.

We previously explored patients' experiences of using WCBT as part of a feasibility trial of the prototype of the WCBT program used in the ACTIB trial [13]. Participants in that study were positive about WCBT and described the website as "well designed and easy to understand and use," although some felt that "a user had to be self-motivated to work through the material." Participants engaged with the website to varying degrees, with some having limited or no engagement because "they did not find the website relevant to them" or "the website was too impersonal." Follow-up was performed at just 12 weeks, so experiences of longer-term effects could not be assessed. To the best of our knowledge, this is the only study on patients' experiences of TCBT and WCBT for IBS. Studies in other populations suggest that WCBT is acceptable and helpful and

allows a level of anonymity when disclosing personal thoughts [14]. Similarly, trials of TCBT in other populations have shown encouraging results in terms of symptom improvement, with no detrimental impact on patient satisfaction [15].

Objectives

We conducted a large qualitative study nested within the ACTIB trial. Previously reported analyses using this data set have focused on treatment seeking and appraisal processes [16], patients' perspectives on general practitioner (GP) interactions [17], and emotional processing in IBS [18]. The aim of this study is to explore patients' experiences and views of TCBT and WCBT for IBS immediately posttreatment and at 12-month follow-up. The objectives are to identify factors that facilitate or impede engagement with web-delivered and telephone-delivered CBT in this patient group both during and after the main intervention period, to identify social and psychological processes of change in the short term and long term, and to provide insight into the quantitative results of this complex trial.

Methods

The ACTIB Trial and Interventions

The ACTIB trial recruited 558 participants from primary care (GP) and secondary care (gastroenterology clinics) in Southampton and London between March 2014 and March 2016. The participants were randomized to one of the following 3 groups: TCBT, WCBT, and TAU. The TCBT group received six 1-hour TCBT sessions over 9 weeks, a detailed patient manual, and 2 *booster* 60-min follow-up phone calls at 4 and 8 months. The WCBT group received access to the previously piloted IBS digital self-management program *Regul8* [4,8]. Regul8 consisted of 8 web-based sessions to be completed on a weekly basis, three 30-min telephone support sessions over 9 weeks, and 2 *booster* 30-min follow-up phone calls at 4 and 8 months. The CBT content delivered via telephone in the TCBT arm and via website in the WCBT arm was the same, with only the mode of delivery being different. Both intervention groups also received ongoing TAU, in primary or secondary care or both, as appropriate. The 2 interventions contained similar content, and the same therapists provided telephone support. The CBT content was based on an empirical cognitive behavioral model of IBS [19] and comprised education and behavioral and cognitive techniques aimed at improving bowel habits, developing stable healthy eating patterns, addressing unhelpful thoughts, managing stress, reducing symptom focusing, and preventing relapse [8]. The TAU group continued with their usual care (in primary care or secondary care or both, as appropriate) and were offered access to Regul8 on completion

of the trial. For further details, see the trial protocol [8]. The study was approved by the relevant National Research Ethics Service Committee on June 11, 2013 (13/SC/0206).

Nested Qualitative Study

Design

A longitudinal qualitative study was nested within the ACTIB trial in an embedded mixed methods design with the qualitative component acting in a supportive capacity [20]. Repeated, also known as serial, semistructured interviews were conducted with the same participants at 3 and 12 months. Serial interviews are rarely used in medical research but, compared with one-off interviews, are better suited to exploring patients' experiences over time and changes therein [21]. Therefore, we chose serial interviews because our objectives were oriented toward processes that occur in time (eg, identifying processes of change in the trial) and because we were interested in how patients'

experiences and reflections might change from the initial therapy phase to the subsequent follow-up phase.

Data Collection

Purposeful sampling was used to select a range of ACTIB participants to invite for interview. To best address our objectives and to capture the experiences of a diverse range of individuals, we sought to interview participants from all 3 ACTIB groups and to include variation within each group in gender, age, ethnic background, geographical location (Southampton or London), symptom severity score, and recruitment path (primary or secondary care). Of 558 participants, 100 were invited to take part in an interview, 58 of whom agreed to participate. The data for this analysis comprised the interviews conducted with people from the TCBT and WCBT groups at 3 months postbaseline (n=34) and at 12 months postbaseline (n=24). The characteristics of these participants are summarized in Table 1, demonstrating the breadth of our sample.

Table 1. Baseline demographic and clinical characteristics of interviewees by trial group.

Characteristics	Therapist CBT[a]		Web-based CBT		Total sample	
	3 months (n=17)	12 months (n=12)	3 months (n=17)	12 months (n=12)	3 months (n=34)	12 months (n=24)
Gender (female), n (%)	13 (76)	10 (83)	14 (82)	9 (75)	27 (79)	19 (79)
Ethnicity, n (%)						
White British	11 (65)	7 (58)	12 (71)	9 (75)	23 (68)	16 (67)
White other	4 (24)	4 (33)	4 (24)	3 (25)	8 (24)	7 (29)
Mixed White and Asian	1 (6)	1 (8)	0	0	1 (3)	1 (4)
African	1 (6)	0	0	0	1 (3)	0
Other ethnicity	0	0	1 (6)	0	1 (3)	0
Age (years), mean (SD)	39.94 (11.71)	38.4 (10.4)	42.41 (17.37)	45 (18.63)	41.18 (14.64)	41.7 (15.14)
Irritable Bowel Syndrome Severity Scoring System baseline score, mean (SD)	283.47 (117.11)	278.58 (126.07)	259.65 (124.39)	219.58 (123.01)	271.56 (119.57)	249.08 (125.35)
Recruitment site, n (%)						
Primary care	11 (65)	8 (67)	13 (76)	9 (75)	24 (71)	17 (71)
Secondary care	6 (35)	4 (33)	4 (24)	3 (25)	10 (29)	7 (29)
Duration of symptoms in years before study entry, mean (SD)	14.71 (7.10)	12.83 (6.94)	15.59 (8.89)	16.33 (9.24)	15.15 (7.93)	14.58 (8.19)
Length of irritable bowel syndrome diagnosis when entering the trial, mean (SD)	7.94 (7.66)	6.5 (6.53)	11.82 (9.22)	13 (8.72)	9.88 (8.59)	9.75 (8.23)

[a]CBT: cognitive behavioral theory.

Interviews were conducted either face-to-face (n=9) or via telephone (n=49), lasted between 22 and 113 min, and were audio-recorded and transcribed verbatim using unique participant identification numbers to preserve anonymity and permit linkage between repeated interviews. A semistructured topic guide was used flexibly, allowing the interviewer to explore any relevant issues raised by the participants. The topic guides for the 3-month and 12-month interviews are available in Multimedia Appendices 1 and 2, respectively. The topic guides included open-ended questions on expectations about the ACTIB trial and reasons for taking part, previous

experiences of IBS therapies and management, experiences of being in the trial and the allocated therapy, and any changes that occurred since starting the trial. Interviews at 3 months continued until data saturation, that is, the point at which no new themes relevant to the research questions were identified. This was reached when 34 participants had been interviewed. The same 34 participants were contacted again at 12 months. Of 34 participants, 24 agreed to be interviewed again. The remaining 10 either failed to respond or declined, citing a lack of time to take part. Data saturation for themes related to posttrial experiences, and longer-term retrospective reflections

on trial experiences was reached within those 24 interviews, making additional recruitment unnecessary.

Data Analysis

Interviews were read repeatedly before being coded in NVivo (QSR International; version 11) and analyzed by working iteratively with the phases mapped out by Braun and Clarke for inductive thematic analysis [22] supplemented with techniques from grounded theory (Table 2) [23,24].

Table 2. Summary of the analytic process.

Thematic analysis phase	Implementation	Supplementary techniques derived from grounded theory
Familiarization	• Initial notes made as transcripts read repeatedly	Listen to audio recordings
Generate initial codes	• Using the first 22 transcripts, initial codes and a coding manual were developed. This coding manual was used to analyze subsequent transcripts, and amendments were made iteratively when necessary	Line-by-line open coding on a portion of the data and constant comparison
Searching for themes	• As the analysis evolved, codes related to similar manifest or latent concepts were grouped together. These groupings were considered as candidate themes and subthemes	Constant comparison, identify key concepts in the data, and write memos
Reviewing themes	• Candidate themes and subthemes were reviewed to ensure that they worked in relation to the coded extracts and the individual interviews and that they captured relevant material from across the data set	Constant comparison, search for deviant cases, generate selected case summaries to capture participant stories, and changes across 3- and 12-month interviews
Defining and naming themes and their interrelations	• Themes were refined and explicitly defined to clearly and succinctly capture patterns in the data relevant to the research objectives • Cross-tabulations (using NVivo's matrix query) to compare theme content and relevance between the TCBT[a] and WCBT[b] groups and between 3 and 12 months	Constant comparison
Reporting	• Selected compelling examples to illustrate themes and subthemes. Final analysis and contextualization in relation to the literature and research objectives	N/A[c]

[a]TCBT: telephone-based cognitive behavioral theory.

[b]WBCT: web-based cognitive behavioral theory.

[c]N/A: not applicable.

An attempt was made to bracket the influence of the researcher's prior knowledge and assumptions on the coding while acknowledging that this is never fully achievable and that the emerging analysis is necessarily a product of the interactions between interviewer, interviewee, and analyst, situated within their particular sociocultural, intellectual, and historical contexts. This analysis was guided by the research objectives, supervised by an investigator experienced in qualitative methods (FB) and led by junior nonclinical researchers trained in qualitative research but not CBT (SH, JH, and AF), one of whom (SH) was very familiar with the Regul8 intervention. They identified initial themes and subthemes, some of which resonated with the theoretical processes underpinning CBT. The initial and final themes and subthemes were reviewed and interpreted (SH, AF, JH, and AS) with input from trained CBT therapists (RM and TC), health psychologists (RM, TC, and FB), and an academic GP (HE). RM and TC led the development of the initial model underpinning the CBT intervention, and HE, SH, AS, and FB were involved in the development of Regul8. The credibility of this qualitative analysis was enhanced by the involvement of multiple researchers and the use of NVivo to facilitate (1) the iterative process of analysis moving between raw data, codes, and themes within a large data corpus and (2) systematic comparisons between and within individual participants.

Results

Overview

A total of 4 main clusters of themes related to each objective were identified and were evident to some extent within both TCBT and WCBT groups: experiencing symptomatic and quality of life improvements; developing a different mindset: cognitive and behavioral changes; barriers to engagement with CBT; and facilitators to engagement with CBT. Each cluster comprised multiple themes, which are summarized in Figure 1. Below, we discuss these in detail, highlighting individual themes in italic typeface.

XSL•FO
RenderX

Figure 1. Summary of themes and subthemes capturing participant experiences of telephone and web-based cognitive behavioral theory for irritable bowel syndrome. CBT: cognitive behavioral theory; IBS: irritable bowel syndrome; WCBT: web-based cognitive behavioral theory.
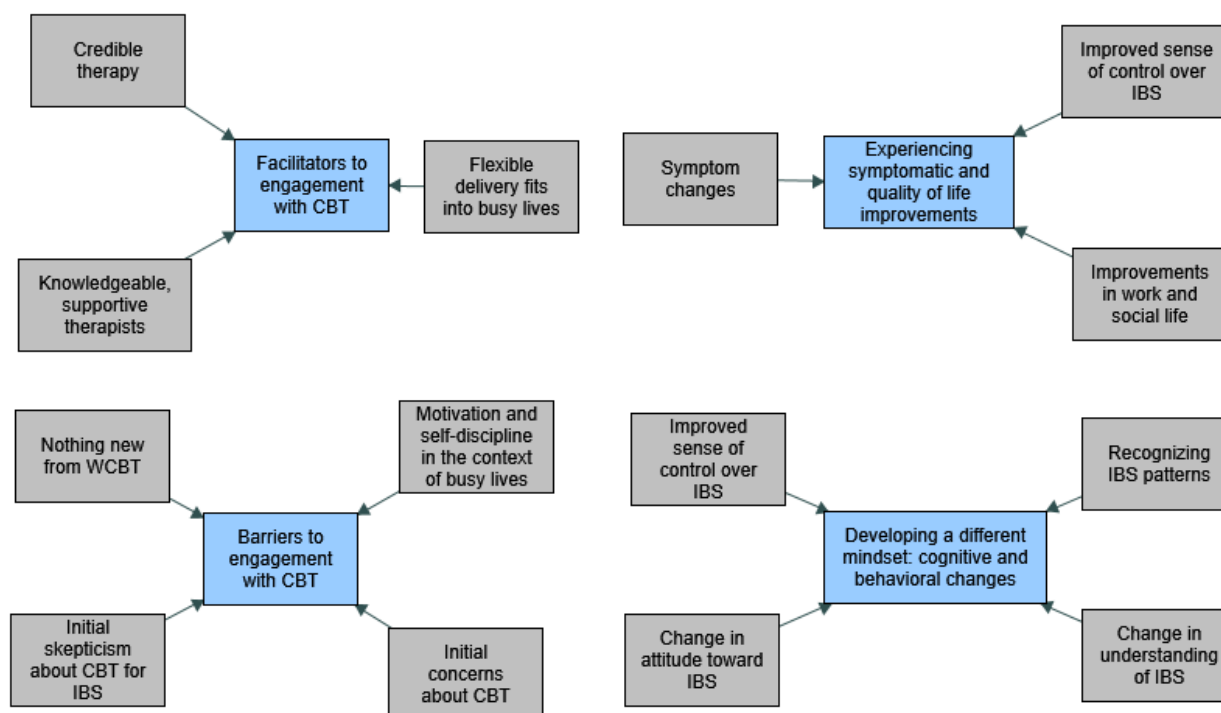
Identifying Factors That Facilitate Engagement With Web-Delivered and Telephone-Delivered CBT for IBS

This objective was addressed by the themes collated under *facilitators to engagement with CBT*. High levels of satisfaction with CBT were suggested by participants' positive comments about their experiences at both 3 months and 12 months and may have either facilitated or reflected high levels of engagement. Participants' views of WCBT and TCBT being credible for IBS were shaped by perceptions that the ACTIB CBT took a systematic well-ordered approach; presented material in a professional, engaging, and accessible manner; and provided clear explanations of IBS and a convincing rationale for CBT. Early improvements in symptoms also contributed to a view of CBT being credible for IBS. This was true for both CBT groups, with no clear differences between the groups. A participant in the TCBT group explained:

> *Literally after the second week of doing it, sort of reading through the books and then—talking to [name of therapist] for the hour and going through everything, it was brilliant and the fact that it did really help, you know, week by week we were talking about different behaviours and—and I think, literally, I sort of saw improvement quite quickly really.* [P24547, 3 months, TCBT]

Similarly, a participant in the WCBT group reported:

> *...overall I was very, very pleased, it was nicely laid out, and I think that it's really contributed into helping me overcome some of the issues that I've dealt with—I've been dealing with—up until that point.* [P10074, 3 months, WCBT]

Participants in both groups valued being able to talk to therapists who were perceived as knowledgeable and supportive. They valued:

> *just having someone to talk to* [P39446, TCBT, 3 months]

and found it particularly helpful to talk to someone who had IBS specific knowledge:

> *it's just nice to have someone listen who is kind of—understands all the ins and outs* [P45322, WCBT, 3 months]

It would seem that the therapists were able to develop positive therapeutic relationships despite being constrained by the telephone to verbal communication only. Affective bonds were also evident, as participants described their therapists as follows:

> *really friendly* [P25044, TCBT, 3 months]
>
> *very nice to speak to* [P20850, TCBT, 3 months]
>
> *friendly and approachable* [P16084, WCBT, 3 months]

One potential drawback of web-based interventions, at least for some patients, is the lack of human interaction and support [12,13]. The telephone sessions provided to support the WCBT mitigated this risk, and participants described how this therapist contact helped to reinforce the messages from Regul8, provided the opportunity to discuss their particular case, and enabled them to ask questions and have them answered. In these ways, the 30-min telephone sessions helped to support engagement with the WCBT, in particular:

> *I think with the actual online sessions, they've been really helpful, but then, being able to talk... to my*

therapist, has really helped me kind of put them into practice. I think, for me, the online sessions themselves weren't enough to deal with my case and my symptoms. I think being able to talk them over with the therapist has kind of—re-cemented things, reaffirmed things and just being able to talk about them has really helped. [P28570, WCBT, 3 months]

talking to [therapist name] was really, really good. That was nice to have that kind of backup support and, you know, I had like—I think it was half an hour and it was a really good amount of time... [P45322, WCBT, 3 months]

The flexibility afforded by both web-based delivery and telephone delivery was commented on positively by participants. Both modes of delivery were experienced as convenient for participants who were able to organize their own schedules for completing the web-based modules and did not have to physically travel to attend appointments:

I liked the way it was presented and it's certainly helped me because I was able to do it my own time. [P16033, WCBT, 12 months]

I liked that it was over the phone sometimes as well, that I didn't always have to go somewhere and park and—that was good. [P25044, TCBT, 12 months]

Identifying Factors That Impede Engagement With Web-Delivered and Telephone-Delivered CBT for IBS

Barriers to engagement with CBT encompassed 4 main themes describing factors that impede engagement with WCBT and TCBT. Participants in both groups who discussed initial skepticism and initial concerns about the biopsychosocial nature of the intervention also reported mostly overcoming these potential barriers to engagement. Some described feeling skeptical about how remote CBT could be either relevant or effective for IBS, and this skepticism was typically couched in a perceived Cartesian disjunction between CBT as focused on mental processes and IBS as a physical condition. Skepticism in both groups was overcome through beginning the intervention and learning about the cognitive behavioral model of IBS used in ACTIB and developed in an earlier trial [25]. The cognitive behavioral model was presented by starting with the physiological and biological changes that underpin the IBS symptoms and then exploring how these changes can be influenced by thoughts, feelings, and behaviors as well as the autonomic nervous system responses linked to stress:

To be honest, when I started I was very skeptical, I couldn't see how thought processes and things would actually affect your tummy, but when it's explained through the literature and when you speak to a therapist, you can really see the connection between how you think and how your tummy reacts and—I think it just takes somebody to tell you... [P25119, TCBT, 3 months]

Some participants across both groups expressed initial concerns about whether the telephone was an appropriate mode of delivery for CBT as a form of talking therapy. For a few (2 TCBT interviewees and 4 WCBT interviewees), these concerns

persisted and appeared to derive from discomfort with the lack of nonverbal cues in telephone interactions:

obviously over the phone it's slightly trickier than face-to-face. So we don't know what the other person's thinking or anything. [P40192, TCBT, 12 months]

Others overcame their concerns once they had started telephone sessions and focused more on the content of the CBT being delivered:

I did think it is odd to do counselling over the phone, but now I think—actually—it doesn't matter, it doesn't really matter at all as long as the counselling is good. [P40210, TCBT, 3 months]

Participants from both the TCBT and WCBT groups referred to the need for self-discipline and motivation to complete the homework tasks contained within the CBT program. Some found it difficult to motivate themselves to do this homework and may have been negatively affected by the connotations of this terminology:

I found it hard um... I'm not very good at doing homework and never have been and I don't suppose I will be, um...so where it's given my homework to do, I've not—I've not been, um let's say a Grade A student. [P21339, WCBT, 3 months]

Participants described wanting to do the homework to experience the anticipated health improvements but also reported finding it difficult to do this within the context of busy lives and competing priorities. There was a sense in both the WCBT and TCBT groups that investing more time in the program would result in getting more out of it:

I: What did you dislike about being in this group?

P: I think probably the discipline of having to do the homework... it's kind of a bit of a paradox; I wanted to do the homework because I'm keen to participate and kind of make the best of it, but it's kind of remembering to do it and—having something else to do during the week. [P25044, TCBT, 3 months]

There could have been more I learned from it if I'd maybe done—spent more time on it or done it over a longer amount of time, then I might have got more out of it, but I did get a lot from it. [P40567, WCBT, 12 months]

One final barrier that impeded engagement with WCBT (but was not identified in the narratives of TCBT participants) for some participants was the sense that this intervention did not offer anything new. This was mostly expressed by participants who felt that they had lived with their IBS for a long time and had already made themselves familiar with and tried to implement recommendations regarding lifestyle issues, including diet, stress, and physical activity. Such participants felt that the WCBT did not offer them any new insights or approaches to managing their IBS:

I've followed all the little sections on the trial, looking at your diet, looking at your stress, looking at your activity and I've kind of gone through all of those on

my own in the last few, you know, over the years. So-from... my point of view, I didn't get an awful lot out of it because it was already telling what I already knew. [P20066, WCBT, 3 months]

Although this theme was only present in the WCBT arm, it is important to note that the content of the CBT program was the same in both TCBT and WCBT arms, with only the mode of delivery being different.

Identifying Social and Psychological Processes of Change

Social and psychological processes of change were captured in the cluster of themes "Developing a different mindset: cognitive and behavioral changes." Within this cluster, 4 themes described different changes experienced by participants across both WCBT and TCBT groups: changes in their understanding of IBS, changed attitudes toward IBS, a newfound ability to recognize IBS patterns, and subsequent changes in behavior associated with an increased sense of control over their IBS. At 3 months, participants in both groups felt they had an improved understanding of IBS as a reassuringly common condition that they could manage to some extent, based on their personal experience of improvements in IBS since commencing the trial:

I'm really pleased [with that] and it doesn't mean the symptoms are completely gone, but it means that I understand them and I can control some of them. So I think that's something I'm going to keep for life; it's not something I'm going to forget about. [P40015, TCBT, 3 months]

I feel relieved, relieved that—because going through the programme I realised that there are people out there who suffer exactly the same symptoms as I do, that it's actually fairly common. [P10074, WCBT, 3 months]

The CBT program taught participants to recognize their personal cognitive and behavioral patterns related to their IBS and enabled participants to evaluate their responses to these patterns. When these ideas resonated with individuals, they were able to reflect on them and think about things in a different way:

I think like with the thoughts, just being aware of the sort of things that can kind of perpetuate the cycle of stress. I think catastrophizing or black-and-white thinking and things like that, I think I can see them in myself and I think just being aware of that, you can kind of try and take a back step and see it in another way and re-evaluate the situation. [P26417, TCBT, 3 months]

Possibly as a consequence of feeling that they understood their IBS better and could identify their personal patterns, participants also reported changing how they thought about IBS. For example, some described being able to feel more relaxed and less worried about their IBS. This change then appeared to be linked with positive changes in behavior, for example, enabling participants to liberate themselves from ingrained and socially limiting behaviors such as avoiding public or shared toilets:

I started thinking to myself, you know, I don't need to stress and worry about it, there is a toilet here, it's

there if I can use it, which has helped, because I used to go home early from parties and things like that with a tummy ache. But now I can just think, you know, I'm working myself up about it and then, more often than not, the feeling goes away and then I'm absolutely fine. [P25119, TCBT, 3 months]

Participants valued the tools and strategies that they developed through the CBT program when they were found to effectively help manage their (response to) IBS. In this way, both WCBT and TCBT appeared to promote an increased sense of control over IBS and greater self-efficacy for coping:

I would say I'm more in control than um previously because I have a whole series of tools to help me. [P20822, TCBT, 3 months]

I feel so much more in control of my IBS and if something flares up I don't feel like it's the end of the world and I know that I've got strategies in place to be able to deal with them. [P28570, WCBT, 3 months]

The cognitive and behavioral changes promoted by the CBT program were summed up by a participant in the TCBT group who experienced total relief from IBS, which was maintained at 12 months:

I now have a different mind-set, if you like, and a few little aids along the way, which help me to—remember the interviews and I don't have any problems and haven't had any problems since the ACTIB course finished. [P20822, TCBT, 12 months]

Insights Into the Quantitative Results of the ACTIB Trial

Cross-tabulating the themes by trial group and interview time point helped to relate the qualitative data to the following key findings from the quantitative trial: the overall effectiveness of WCBT and TCBT, the maintenance of effects at 12 months, and adherence to the interventions.

In the trial, all primary and secondary outcomes showed significant improvement in both CBT groups compared with TAU at 12 months [9] (the primary end point), and the overall pattern was for beneficial effects (in IBS symptom severity, mood, and impact on life roles) to be sustained, on average, from 3 to 6 to 12 months. There was also evidence of some sustained improvements and some attenuation of effects from 12 to 24 months [10]. The theme *Experiencing symptomatic and quality of life improvements* captures the range of patient-perceived benefits of the CBT program. The majority of participants in both therapy arms reported IBS symptom improvements over the period of the study; some participants also described how the symptomatic, cognitive, and behavioral changes associated with CBT contributed to improvements in their work and social lives. These improvements were particularly emphasized when participants were interviewed at 3 months but were still very evident in the 12-month interviews, demonstrating the lasting positive impact on the lives of these IBS patients:

I think it is much improved really; I've not had as much sort of constipation as what I used to have,

so—so yes—for me, it has been really, really good. [P24547, TCBT, 3 months]

if I do have a problem I know I can, you know, let my boss know and she's fine. I'll just say—I'm going to be in a bit late, whereas, you know, like I said before, I just—I would have just been like—I'm sick. [Pt 45322, WCBT, 3 months]

I used to be very concerned about going round to people's houses or going out to dinner—or going for food somewhere because I'd get very concerned that I might have a reaction and I'd need to run to the toilet straightaway, and I think that stress and worry beforehand would always then trigger a bout of IBS, but now it's kind of—I started thinking to myself, you know, I don't need to stress and worry about it, there is a toilet here, it's there if I can use it, which has helped. [Pt 25119, TCBT, 3 months]

However, while some participants maintained improvements and felt they would carry these on into the future, others felt they had not managed to maintain earlier improvements:

Unfortunately I'm unable to—control it as I did when I did the study and even though I still do a number of techniques and use the tools that I have learned last year, I've come to think that my mind has become immune to them and knows that these are just things I'm telling myself, but it's not registering; so the mind controls my body, still. [P10074, WCBT, 12 months]

To explore possible reasons why some participants felt they did not sustain beneficial changes after completion of active treatment, we classified participants as responders (n=22) and nonresponders (n=12) at 12 months and compared the themes and subthemes across these groups. A *responder* was defined as a participant with a 50-point improvement on the Irritable Bowel Syndrome Severity Scoring System from baseline to 12 months [8]. This analysis suggested that people classified as responders had more positive experiences of active treatment than those classified as nonresponders. At 3 months, the responders talked more about developing a different mindset and making cognitive and behavioral changes in response to CBT. Nonresponders placed more emphasis on barriers to engagement. This suggests that patients' engagement with structured active therapy and their ability to embed cognitive and behavioral changes in their lives are, unsurprisingly, important for longer-term effects. Two case summaries presented in Textbox 1 help to illustrate this within the broader context of patients' experiences. Participants' reflections also suggest that any attenuation of beneficial effects could be partially mitigated by providing limited ongoing access to a therapist to help discourage relapse into unhelpful patterns:

I would have found it very useful, as I'm sure most participants would, to perhaps long-term, for this kind of treatment, have maybe somebody that you could contact, a point of contact from time to time, when you were having a particularly difficult time or needed to be reminded of something or to re-motivate you because like with all things, if we don't have somebody behind us, I think we tend to have good intentions and then just go back to our old habits. [P20850, TCBT, 12 months]

Textbox 1. Case summaries of responders and nonresponders to telephone-based cognitive behavioral therapy.

Two participants were selected for in-depth presentation to illustrate how one responder and one nonresponder experienced cognitive behavioral therapy (CBT) for irritable bowel syndrome (IBS). These are not presented as representative cases but rather to showcase the interplay between themes as participants answered open-ended interview questions about their experiences over the course of a CBT trial. Both started the trial with *severe* IBS symptom severity scores, both were randomized to the telephone-based CBT arm, and both completed all of their telephone sessions including their booster calls. Participant P24547 maintained improvements at 12 months and was classified as a responder, whereas P40210 did not.

- Participant P24547

P24547 had experienced IBS symptoms for 15 years and did not really expect to benefit personally from the trial, having previously tried many treatments and "sort of thought it was something I just had to learn to live with, I didn't think that I'd get a huge amount out of [the trial]." She had not previously tried CBT and did not express any reservations about it. When describing the nature of CBT and its impact on her IBS, P24547 emphasized the cognitive aspects of the treatment. For example, when asked about any changes experienced since starting the trial, she described how "I used to get quite stressed and worked up about—people at work. So [therapist name] was always saying that a lot of people with IBS have got sort of a level of perfection in themselves and I've definitely changed in the fact that I—I'm trying to do as much as I can, the best—but I don't always—the best I can but I don't always strive for perfection—which—which I think is that sort of side me has really helped, working through those exercises." Related to this perceived need to make active changes to one's thinking patterns, a strong sense of empowerment to make such changes and thus manage any symptoms emerged from P24547's account of CBT. "I think through this therapy it really sort of highlighted some of the things, even simple things like doing more exercises and—making myself go out and—and not stay in and just sort of think and worry about my pain and get annoyed because of why this is happening to me, sort of thing. It was sort of understanding that there are things that I can do to work with the idea. So—for me—it was sort of being able to recognize the symptoms and then know how to deal with them." During the trial, P24547 found the interactions with the therapist helped to motivate her to practice implementing her new cognitive and behavioral styles and was concerned that she might struggle to sustain these changes after completing the trial, but at 12 months explained how "actually I have continued to reflect on things and if I get cross about something, instead of getting myself wound up, which then tends to make my IBS even worse, I do—even literally last week—I got really cross about how something went and I then thought, no; I went to the toilet, I breathed, and I thought, right, how can I see this from their point of view, which is what the manual often went through. So even now I'm still finding it really useful."

- Participant P40210

P40210 had experienced IBS symptoms for 24 years and was struggling with her IBS symptoms, feeling desperate for help at the start of the trial and "prepared to just try anything, I would have been happy to stand on my head if it had made it better." She had "done CBT before for depression" and was a little apprehensive about delving into her emotions to start with but appreciated the new insights that she gained from CBT for IBS "now I know they [thoughts and feelings] are totally linked to my problems. So that's been a positive thing that's come out of it." When asked about the effects of CBT, she evidenced her new insight into her IBS triggers but emphasized the improvements in her IBS symptoms and did not describe having adopted any cognitive or behavioral changes. "I have—what symptoms I'm having now ... are a huge amount less, I mean a massive amount less. So it's mainly now wind, a bit of rumbling tummy and a little bit of—a little bit of acid reflux, a very small amount, but I can actually control that if I don't eat certain things. And also—I have had one incident—a sort of very small amount of soiling incident, which is very unusual, I haven't had that for about—three or four years. That was a day when I was particularly very stressed. And—small amounts of pain, you know, sort of spasm-type pain, sort of low down in my tummy, but really—a huge amount less than I had before." P40210's account of CBT did not suggest that she thought it was necessary to actively work on making cognitive changes to help her IBS. Instead, she seemed to engage with CBT on a more limited basis, accepting a new understanding of IBS but not acting on that understanding, instead focusing on dietary measures. Without cementing the underpinning cognitive changes, P40210 did not manage to sustain her dietary changes at 12 months: "But now a year later, I think I've probably fallen back into probably old habits really and I think a lot of that is to do with the fact that—and I was discussing this with my daughters earlier—the fact that my IBS problems have been going for a considerable length of time. So it's as if the problems have outweighed the solutions, you know, the problems are more dominant than the solutions."

Quantitative data suggested good adherence to TCBT and WCBT. In TCBT, 83.9% (156/186) of participants completed at least four telephone sessions, 88.1% (163/185) completed at least one telephone session, and 69.2% (128/185) completed at least four website sessions. Previous trials have varied in the way they have defined *adherence*, and reaching an appropriate definition is not straight forward [26]. We predefined *adherence* in our protocol and based it on the notion that participants would need to have received at least half of the program to be considered to have received CBT. Only 3 interviewees were classified as nonadherent in our trial, all of whom had received WCBT, and we were unable to identify any themes that differentiated adherent and nonadherent participants.

Discussion

Principal Findings

Participants provided positive feedback about both web-based and telephone-based CBT for IBS. They described

improvements in IBS symptoms, positive changes in their understanding of and attitude toward IBS, and a newfound ability to recognize IBS patterns and change their own behaviors. This resulted in an increased feeling of control over their IBS and improved work and social life despite some initial skepticism regarding remote CBT for IBS. They highlighted the need for self-discipline to undertake CBT and maintain behavioral changes in the longer term but felt that the flexibility of telephone-based or web-based CBT and high-quality therapy input aided engagement. Telephone support in the WCBT group was important and valuable to the participants.

Strengths and Limitations

This study was a rigorously conducted qualitative study that benefited from interviewing participants just after the CBT interventions to gather immediate perceptions at the end of the treatment (3 months) along with longer-term perceptions at 12 months. Participants were not interviewed again at 24 months, which reduced our ability to relate our qualitative findings to

the 24-month quantitative follow-up [10]. The qualitative interviews enabled an exploration of individual differences in responding that were masked by the necessary focus in the trial data on group-level differences and changes over time. Participants were recruited from both primary and secondary care, which encompassed participants at different stages of their IBS journey, improving transferability to different settings. Participants interviewed had volunteered to participate, and all barring 3 were classified as *adherent* to the program. The results may have differed from a non–self-selecting, nonadherent sample. Participants were mostly White, British, and female, which is representative of the main ACTIB trial sample, but this sample does not allow us to draw inferences about how people from other demographic groups might experience CBT for IBS.

Although various measures were taken to minimize any inappropriate or uncritical influence of investigators' preexisting frameworks, it is important to acknowledge that the interpretation of the data may have differed if conducted by a different research team. The measures taken to enhance analytic rigor included coding of data conducted by 3 different individuals, obtaining input throughout the analysis process from the multidisciplinary team, and consciously exploring possible alternate explanations and discussing them within the team. Furthermore, the individuals coding the data were not CBT therapists and did not consult the CBT therapists or model underpinning the intervention until the interpretation phase, after the subthemes had emerged.

Comparison With Existing Literature

To the authors' knowledge, this is the largest qualitative study to date exploring the experiences of participants undertaking web-based and telephone-based CBT for IBS. This study expands on work by Tonkin-Crine et al [13] by:

- including participants at 2 time points: 12 weeks and 12 months postbaseline, rather than just 12 weeks
- using a larger sample
- including participants who had received telephone-based CBT, rather than just web-based CBT

The results from this study showed similarities to the findings from the study by Tonkin-Crine et al [13], for example, the feelings of positivity about WCBT, and the need for self-motivation to carry them through the program. However, not all findings were replicated, notably, the current findings did not describe the website as *impersonal* or not personally *relevant enough*. This difference in findings may be accounted for by the different levels of telephone support in each study; WCBT participants in this study received three 30-min telephone sessions and two 30-min booster sessions, all with a trained CBT therapist. WCBT participants in the study by Tonkin-Crine et al [13] received much less telephone contact (1 session of 30-45 min), which was conducted by a practice nurse. Participants in this study valued the telephone support and the expert knowledge of the therapists, and it may be that this extra contact time with expertly trained therapists addressed these

previously reported barriers. Indeed, having individuals providing support who are perceived as trustworthy, benevolent experts may be vital for such support to effectively engage patients in digital interventions [27].

Conclusions and Implications for Future Research

Although acknowledging the difference in adherence rates between the TCBT and WCBT groups, more research is needed to investigate ways of increasing engagement in the web intervention. This may include a therapist portal where therapists can see what patients have completed on the web and provide support and encouragement online as well as by telephone. The data showed the telephone support in the WCBT group to be valued and important, and keeping this element in any future version would be beneficial.

Future versions of this CBT program may benefit from addressing the identified barrier around patients who have had IBS for a long-suffering feeling the web program *does not offer anything new*. It may be helpful for therapists to use Socratic questioning or guided discovery when discussing the patients' personal cognitive behavioral model in the initial telephone call to unpick familiar and unfamiliar areas and, if appropriate, provide reassurance that more novel content will be covered later in the program.

Future CBT programs for IBS might benefit from addressing potential skepticism about the effectiveness of such treatment at the start of the program to help those with IBS understand how it might help. It is important to note that although this study focused on remote CBT for IBS, skepticism may be applicable more generally to CBT for IBS, rather than the delivery mode.

It is unclear why the theme around the website failing to *offer anything new* was only present in the WCBT arm; however, perhaps, the therapists in the TCBT arm were able to better personalize the content and focus on the parts most novel or applicable to the participants. Participants in the WCBT group indicated that the small amount of telephone support they received was helpful to keep them on track and provided an outlet to ask questions and talk about their progress. Future research may investigate the minimal amount of therapist contact time needed in the delivery of an effective web-based program for IBS, which may fall somewhere between 30 and 45 min of nurse contact offered in the previous trial [13] and the 150 min of therapist contact offered in the ACTIB trial. In addition, the longevity of this support and contact time needs to be explored, as some participants expressed the desire for longer-term support they could return to when they started to slip back into old habits. The potential source of this longer-term support also needs to be explored, for example, it is unclear whether this support would need to be provided by a therapist, a website, or a patient's GP.

Both TCBT and WCBT for IBS were positively received by people with refractory IBS. The flexibility and perceived high quality of the interventions aided engagement. These forms of CBT have the potential to provide a lower-cost acceptable alternative to face-to-face CBT.

Conflicts of Interest

Multimedia Appendix 1
Three-month interview topic guide.
[DOCX File , 25 KB - jmir_v22i11e18691_app1.docx]

Multimedia Appendix 2
Twelve-month interview topic guide.
[DOCX File , 25 KB - jmir_v22i11e18691_app2.docx]

References

1. Irritable Bowel Syndrome in Adults: Diagnosis and Management. National Institute for Health and Care Excellence. 2017. URL: https://www.nice.org.uk/guidance/cg61/resources/irritable-bowelsyndrome-in-adults-diagnosis-and-management-975562917829 [accessed 2020-09-18]

2. Hunt MG, Moshier S, Milonova M. Brief cognitive-behavioral internet therapy for irritable bowel syndrome. Behav Res Ther 2009 Sep;47(9):797-802. [doi: 10.1016/j.brat.2009.05.002] [Medline: 19570525]

3. Ford AC, Quigley EM, Lacy BE, Lembo AJ, Saito YA, Schiller LR, et al. Effect of antidepressants and psychological therapies, including hypnotherapy, in irritable bowel syndrome: systematic review and meta-analysis. Am J Gastroenterol 2014 Sep;109(9):1350-65; quiz 1366. [doi: 10.1038/ajg.2014.148] [Medline: 24935275]

4. Everitt H, Moss-Morris R, Sibelli A, Tapp L, Coleman N, Yardley L, et al. Management of irritable bowel syndrome in primary care: feasibility randomised controlled trial of mebeverine, methylcellulose, placebo and a patient self-management cognitive behavioural therapy website. (MIBS trial). BMC Gastroenterol 2010 Nov 18;10:136 [FREE Full text] [doi: 10.1186/1471-230X-10-136] [Medline: 21087463]

5. Lackner JM, Jaccard J, Keefer L, Brenner DM, Firth RS, Gudleski GD, et al. Improvement in gastrointestinal symptoms after cognitive behavior therapy for refractory irritable bowel syndrome. Gastroenterology 2018 Jul;155(1):47-57 [FREE Full text] [doi: 10.1053/j.gastro.2018.03.063] [Medline: 29702118]

6. Zijdenbos I, de Wit NJ, van der Heijden GJ, Rubin G, Quartero A. Psychological treatments for the management of irritable bowel syndrome. Cochrane Database Syst Rev 2009 Jan 21(1):CD006442. [doi: 10.1002/14651858.CD006442.pub2] [Medline: 19160286]

XSL•FO
RenderX

7. Kennedy T, Jones R, Darnley S, Seed P, Wessely S, Chalder T. Cognitive behaviour therapy in addition to antispasmodic treatment for irritable bowel syndrome in primary care: randomised controlled trial. Br Med J 2005 Aug 20;331(7514):435 [FREE Full text] [doi: 10.1136/bmj.38545.505764.06] [Medline: 16093252]

8. Everitt H, Landau S, Little P, Bishop FL, McCrone P, O'Reilly G, ACTIB trial team. Assessing cognitive behavioural therapy in irritable bowel (ACTIB): protocol for a randomised controlled trial of clinical-effectiveness and cost-effectiveness of therapist delivered cognitive behavioural therapy and web-based self-management in irritable bowel syndrome in adults. BMJ Open 2015 Jul 15;5(7):e008622 [FREE Full text] [doi: 10.1136/bmjopen-2015-008622] [Medline: 26179651]

9. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, ACTIB Trial Group. Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. Gut 2019 Sep;68(9):1613-1623 [FREE Full text] [doi: 10.1136/gutjnl-2018-317805] [Medline: 30971419]

10. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, et al. Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial. Lancet Gastroenterol Hepatol 2019 Nov;4(11):863-872 [FREE Full text] [doi: 10.1016/S2468-1253(19)30243-2] [Medline: 31492643]

11. Francis C, Morris J, Whorwell P. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 Apr;11(2):395-402 [FREE Full text] [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]

12. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010 Dec 7;8:82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]

13. Tonkin-Crine S, Bishop FL, Ellis M, Moss-Morris R, Everitt H. Exploring patients' views of a cognitive behavioral therapy-based website for the self-management of irritable bowel syndrome symptoms. J Med Internet Res 2013 Sep 3;15(9):e190 [FREE Full text] [doi: 10.2196/jmir.2672] [Medline: 24001787]

14. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]

15. Burgess M, Andiappan M, Chalder T. Cognitive behaviour therapy for chronic fatigue syndrome in adults: face to face versus telephone treatment: a randomized controlled trial. Behav Cogn Psychother 2012 Mar;40(2):175-191. [doi: 10.1017/S1352465811000543] [Medline: 21929831]

16. Harvey JM, Sibelli A, Chalder T, Everitt H, Moss-Morris R, Bishop FL. Desperately seeking a cure: treatment seeking and appraisal in irritable bowel syndrome. Br J Health Psychol 2018 Sep;23(3):561-579 [FREE Full text] [doi: 10.1111/bjhp.12304] [Medline: 29508539]

17. Sibelli A, Moss-Morris R, Chalder T, Bishop FL, Windgassen S, Everitt H. Patients' perspectives on GP interactions after cognitive behavioural therapy for refractory IBS: a qualitative study in UK primary and secondary care. Br J Gen Pract 2018 Sep;68(674):e654-e662 [FREE Full text] [doi: 10.3399/bjgp18X698321] [Medline: 30061195]

18. Sibelli A, Chalder T, Everitt H, Workman P, Bishop FL, Moss-Morris R. The role of high expectations of self and social desirability in emotional processing in individuals with irritable bowel syndrome: a qualitative study. Br J Health Psychol 2017 Nov;22(4):737-762. [doi: 10.1111/bjhp.12264] [Medline: 28862389]

19. Spence MJ, Moss-Morris R. The cognitive behavioural model of irritable bowel syndrome: a prospective investigation of patients with gastroenteritis. Gut 2007 Aug;56(8):1066-1071 [FREE Full text] [doi: 10.1136/gut.2006.108811] [Medline: 17324974]

20. Creswell J, Plano CV. Designing and Conducting Mixed Methods Research. Thousand Oaks, California: Sage; 2011.

21. Murray SA, Kendall M, Carduff E, Worth A, Harris FM, Lloyd A, et al. Use of serial qualitative interviews to understand patients' evolving experiences and needs. Br Med J 2009 Sep 28;339:b3702. [doi: 10.1136/bmj.b3702] [Medline: 19786485]

22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

23. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications; 1990.

24. Glaser BG, Strauss A. The Discovery of Grounded Theory: Strategies for Qualitative Research. New Brunswick (USA) and London (UK): Aldine; 1967.

25. Moss-Morris R, McAlpine L, Didsbury LP, Spence MJ. A randomized controlled trial of a cognitive behavioural therapy-based self-management intervention for irritable bowel syndrome in primary care. Psychol Med 2010 Jan;40(1):85-94. [doi: 10.1017/S0033291709990195] [Medline: 19531276]

26. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

27. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

Abbreviations

ACTIB: Assessing Cognitive behavioral Therapy in Irritable Bowel
CBT: cognitive behavioral theory
GP: general practitioner
IAPT: Improving Access to Psychological Therapies
IBS: irritable bowel syndrome
NHS: National Health Service
NIHR: National Institute for Health Research
TAU: treatment as usual
TCBT: telephone-based cognitive behavioral theory
WBCT: web-based cognitive behavioral theory

XSL•FO
RenderX

Original Paper

Engagement With a Web-Based Intervention to Reduce Harmful Drinking: Secondary Analysis of a Randomized Controlled Trial

Paul U Nordholt[1], Dipl-Psych; Eva Christalle[2], MSc; Jördis M Zill[2*], Dipl-Psych, PhD; Jörg Dirmaier[2*], Dipl-Psych, PD, PhD

[1]Department of Nursing and Management, Faculty of Business and Social Sciences, Hamburg University of Applied Sciences, Hamburg, Germany

[2]Institute and Outpatient Clinic of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[*]these authors contributed equally

Corresponding Author:
Jörg Dirmaier, Dipl-Psych, PD, PhD
Institute and Outpatient Clinic of Medical Psychology
University Medical Center Hamburg-Eppendorf
Department of Medical Psychology
Martinistr 52
Hamburg, 20246
Germany
Phone: 49 407410 ext 59137
Email: dirmaier@uke.de

Abstract

Background: Engagement with digital behavior change interventions (DBCIs) is considered a prerequisite for intervention efficacy. However, in many trials on DBCIs, participants use the intervention either only little or not at all.

Objective: To analyze engagement with a web-based intervention to reduce harmful drinking, we explored (1) whether engagement with a web-based alcohol intervention is related to drinking outcomes, (2) which user characteristics are associated with measures of engagement, and (3) whether reported outcomes are associated with data captured by voluntary intervention questionnaires.

Methods: We analyzed data of the intervention arm of a randomized controlled trial on a DBCI to reduce risky alcohol consumption. Data were collected at baseline (T0), after 90 days (T1), and at the end of the 180-day usage period (T2). Engagement with the intervention was measured via system usage data as well as self-reported usage. Drinking behavior was measured as average daily alcohol consumption as well as the number of binge drinking days. User characteristics included demographics, baseline drinking behavior, readiness to change, alcohol-related outcome expectancies, and alcohol abstinence self-efficacy. Following a bivariate approach, we performed two-tailed Welch's *t* tests and Wilcoxon signed rank/Mann-Whitney *U* tests or calculated correlation coefficients.

Results: The data of 306 users were analyzed. Time spent engaging with the intervention as measured by system usage did not match self-reported usage. Higher self-reported usage was associated with higher reductions in average daily alcohol consumption (T1: ρ=0.39, *P*<.001; T2: ρ=0.29, *P*=.015) and in binge drinking days (T1: ρ=0.62, *P*<.001; T2: ρ=0.3, *P*=.006). Higher usage was reported from users who were single (T1: *P*<.001; T2: *P*<.001), users without children (T1: *P*<.001; T2: *P*<.001), users who did not start or finish secondary education (T1: *P*<.001; T2: *P*<.001), users without academic education (T1: *P*<.001; T2: *P*<.001), and those who worked (T1: *P*=.001; T2: *P*=.004). Relationships between self-reported usage and clinical or psychological baseline characteristics were complex. For system usage, the findings were mixed. Reductions in drinking captured by intervention questionnaires were associated with reported outcomes.

Conclusions: Though self-reported usage could be consistently linked to better outcomes and multiple user characteristics, our findings add to the overall inconclusive evidence that can be found throughout the literature. Our findings indicate potential benefits of self-reports as measures of engagement and intervention questionnaires as a basis for tailoring of intervention content. Future studies should adopt a theory-driven approach to engagement research utilizing psychometrically sound self-report questionnaires and include short ecological momentary assessments within the DBCIs.

Trial Registration: German Clinical Trials Register DRKS00006104; https://tinyurl.com/y22oc5jo

XSL·FO

RenderX

Introduction

In many trials on self-guided digital behavior change interventions (DBCIs), a substantial proportion of participants use the intervention only little or not at all [1]. This phenomenon might limit the validity of trial results, since some form of engagement with the intervention is assumed to be a condition for its efficacy [2]. Therefore, it is considered good practice to describe how engagement was measured and report usage along with the primary outcomes when conducting a randomized controlled trial (RCT) on a DBCI [3].

Most likely due to the interdisciplinarity of the field, despite efforts to find a common language [4], the terms engagement, adherence, dose, and usage are often used interchangeably in the literature, and different theoretical perspectives on the topic have emerged [5]. Perski et al [6] found that the behavioral sciences focus on quantitative aspects of engagement, conceptualizing it as the amount of usage over time. However, in computer science literature, research on engagement with DBCIs focuses on its qualitative aspects. In defining engagement as "[…] (1) the extent (eg, amount, frequency, duration, depth) of usage and (2) a subjective experience characterized by attention, interest, and affect," Perski et al [6] highlight the importance of both aspects to obtain a better understanding of engagement and its relation to behavior change. Furthermore, Yardley et al [1] argue that mere engagement with the intervention content is not sufficient to bring about behavior change. Rather, after a phase of engagement with the intervention content, individuals must actively engage in behavior change to achieve the desired outcomes.

In the majority of studies on eHealth and mobile health interventions, engagement is operationalized as the actual usage of the intervention and is measured based on system usage data (SUD) [7]. SUD includes but is not limited to the number of logins, number of modules completed, and the amount of time spent in the intervention. While these measures seem to be predictive of outcomes for interventions targeting physical health, findings for mental health interventions are mixed [8]. Concerning interventions aiming to reduce harmful alcohol consumption, there is some evidence for engagement-outcome relationships [9,10]. However, some studies could not find an association between engagement and the extent of alcohol reduction [11]. It has been argued that the overall inconclusive findings might be due to a lack of consensus on how to measure engagement [8] or different patterns of usage [7]. The latter claim is supported by Milward et al [12] who found that some users of a smartphone app aimed at reducing harmful drinking used it to track their alcohol consumption, while others also used its other features. Furthermore, individuals might differ with respect to the extent of intervention usage required for them to engage in behavior change [1]. Discontinuing the use of an intervention might therefore not be necessarily linked to worse outcomes. It can also mean that the individual's information needs are satisfied and they are already engaged in changing their behavior. A person showing high intervention usage, on the other hand, might have trouble transferring intervention content to actual behavior change and may keep on looking for support in the intervention content. SUD can only capture aspects of the micro level of engagement with the intervention; however, the macro level of engagement with behavior change is a necessity for improved outcomes. Therefore, Yardley et al [1] recommend the utilization of multiple measures of engagement. A cost-efficient way to capture different aspects of engagement are self-report questionnaires. However, at this point, only few of such questionnaires are available and have been psychometrically evaluated [7,13].

As noted above, some form of engagement is necessary to achieve behavior change. To further increase the reach and impact of DBCIs, it is important to identify factors that influence engagement. To date, various user characteristics have been proposed to be related to engagement with intervention content, and overall, the evidence is puzzling [14]. A systematic review identified female gender and a higher treatment expectancy as predictive of higher engagement. Findings for age and baseline problem severity were mixed [15]. Concerning DBCIs targeting alcohol consumption, higher age, higher education, and female gender were found to predict higher engagement [11,16,17]. For baseline alcohol consumption, positive and negative associations with engagement have been found, while some studies could not establish such a relationship [11,17]. Concerning psychological variables, higher motivation for treatment and behavior change as well as lower confidence in one's own ability to change were found to be associated with higher engagement [16,17].

To improve engagement and consequently intervention outcomes, different approaches have been suggested. One of them is tailoring. Tailoring seems to be a desired and efficient way to increase engagement for different types of interventions [15,18,19]. Dynamic tailoring of intervention content relies on the user providing information at some point of the interaction with the DBCI. Based on this user input, the personalized content can be delivered automatically by the system. Apart from static variables such as age or gender, it seems reasonable to consider changes in the target behavior as a basis for tailoring of content. Based on such data, those users who have already changed their behavior could receive content concerning relapse prevention. For others who are still struggling, additional content on behavior change techniques or motivational content could be provided. However, to our knowledge, it is yet unclear whether data captured by intervention questionnaires corresponds to reported outcomes.

Based on data collected between 2015 and 2017 in an RCT on a web-based intervention to reduce risky alcohol consumption—Vorvida [20]—we analyzed engagement with

the intervention. To do so, we explored the following 3 questions.

Engagement and outcomes: How is engagement with the intervention related to drinking outcomes?

User characteristics and engagement: Which user demographics and clinical or psychological baseline characteristics are associated with measures of engagement?

Intervention questionnaires and outcomes: Are reported outcomes associated with data captured by voluntary intervention questionnaires?

As much as possible, post hoc analyses were guided by the AMUsED (Analyzing and Measuring Usage and Engagement Data) Framework [21].

Methods

RCT Details

For detailed information on the study design and procedure of the RCT as well as the intervention, see the publication of the results [20] and the study protocol for the RCT [22]. The RCT was registered at the German Clinical Trials Register (reference number: DRKS00006104), and ethical approval was obtained from the Ethics Committee of the Hamburg State Chamber of Physicians (reference number: PV4802).

Study Design and Procedure

The data we explored in this study were collected from the intervention group of a parallel group pragmatic RCT at 3 time points: at baseline (T0) as well as at 3 months (T1) and 6 months (T2) later. Because access to the intervention was granted for a period of 180 days, T1 corresponds to the middle and T2 corresponds to the end of the usage period. Participants were recruited via health care providers (eg, outpatient centers, information centers, family doctor's offices) in northern Germany and German online and offline media (internet self-help forums, newspaper advertisements). Inclusion criteria were informed consent, a minimum age of 18 years, and either an average consumption of at least 12/24 g (women/men) of pure alcohol per day or an AUDIT-C (Alcohol Use Disorders Identification Test-Concise) score of at least 3 or both.

After filling out a screening questionnaire at the study website, within 1 week, participants received an email informing them whether inclusion criteria were met. Included participants received a link to the baseline questionnaire and were asked to respond. Right after baseline assessment, participants were randomly allocated to either the intervention or the care as usual/waitlist group. A centralized, software-driven, computerized, simple randomization procedure was used so that randomization could not be subverted by the team of researchers and concealed allocation was ensured. Participants in the intervention group received an access key and a link to the intervention. After registration, the program was activated for 180 days. To encourage usage, participants in the intervention group received an email reminder at 3, 6, 9, and 16 weeks after receiving the access key. Six months after the T2 assessment, participants in the care as usual/waitlist group gained access to the intervention.

Intervention

Vorvida is an unguided web-based intervention, which was designed for persons older than 18 years who consider their alcohol consumption patterns to be problematic. The intervention content is tailored to the user depending upon his/her answers in simulated dialogues. Vorvida is based primarily on principles and techniques used in cognitive behavioral therapy. It incorporates different techniques to change behavior and cognition, eg, motivational interviewing, cognitive restructuring, or mindfulness-based methods.

Vorvida is organized in 4 modules. The first module is concerned with individual drinking patterns, motivation to change, and goal setting. The second module focuses on cognitive and mindfulness-based methods to cope with alcohol cravings. The third module informs about coping with risk situations, while the fourth module content is based on relapse prevention. Three short questionnaires are embedded within the program, and users are encouraged to respond to these regularly (self-monitoring). Vorvida can be accessed via standard internet browsers on any desktop, laptop, or mobile computers, including smartphones. Font and image sizes adjust automatically to different screen sizes (responsive design). In addition, Vorvida offers homework materials in the form of PDFs to print and use outside of the intervention.

Measures

Details

In the following section, we briefly describe the measures used, including the abbreviations we will use to refer to them. An overview of all measures used in the analyses, including a short description as well as information on when the assessment took place and references to additional information [23-27], are given in Multimedia Appendix 1.

Drinking Behavior

Drinking behavior was measured in 2 ways: first, as the average daily alcohol consumption in grams of alcohol within the last 7 days by the Timeline Followback (TFB) method [23]; second, as the number of binge drinking days (BDDs) within the last 30 days by asking on how many days the participants drank 5 or more alcoholic beverages on 1 occasion. For research questions (1) and (3), we calculated the individual change scores.

Engagement

Engagement was operationalized as usage of the intervention and measured by SUD and self-report. SUD included the frequency of the utilization of the intervention questionnaires and the time spent in the intervention measured as five-minute blocks of activity (FMB). The FMB was calculated based on server logs. To avoid overestimation of time spent in the intervention, if a user did not create an event (eg, opening a new page) within a time-frame of 5 minutes, he/she was considered inactive. Three different questionnaires were included in the intervention: a daily alcohol check (DAC), assessing the number of standard units of alcoholic drinks consumed on the day before, a weekly alcohol check (WAC) based on the AUDIT-C [28], and "a mood check" (MC) (affective checklist). As an indicator for the breadth of usage, we included the frequency

of use of these questionnaires in our analyses. All SUD were measured continuously and were provided by the GAIA AG—the company responsible for the development and deployment of Vorvida. Self-reported usage (SRU) was assessed at T1 and T2 by asking the participants to rate on a 6-point scale how often they used Vorvida in the past 3 months.

Sociodemographic Data

Sociodemographic data were collected at T0 and these data included the participants' age, gender (female/male), relationship status (single/nonsingle), children (no children/one or more children), education (general higher education entrance qualification [Abitur]/no general higher education entrance qualification [no Abitur]), job qualification (academic studies/no academic studies), and job status (working/not working).

Baseline User Characteristics

Baseline user characteristics consisted of psychological constructs as well as clinical measures concerning drinking behavior. Clinical measures were the age of the first alcohol consumption, the age of regular alcohol consumption, average daily alcohol consumption in grams of alcohol, number of BDDs, and the number of drinking days [23] at baseline (T0). Additionally, we assessed the motivation to change with a readiness ruler [24] and readiness to change with the Readiness to Change Questionnaire [27]. Alcohol abstinence self-efficacy was measured using the Alcohol Abstinence Self-Efficacy Scale [25] and alcohol expectancies were assessed with the Comprehensive Alcohol Expectancy Questionnaire [26].

Data Preparation

Since we were interested in the relationship between usage and changes in drinking behavior, we calculated individual change scores for reductions in the TFB and BDD between T0 and T1 (T1) and between T1 and T2 (T2) as well as total reductions (TOT). We calculated the frequency of utilization of the intervention questionnaires and the time spent in the intervention for the same intervals. To determine whether changes in the DAC or WAC predicted changes in the drinking behavior at T1 or T2, we calculated values for these questionnaires as follows: (1) baseline value: DAC, mean of the individual scores within 14 days starting with the first utilization, and WAC, individual

score of the first utilization; and (2) 30-day value/90-day value: DAC, mean of the individual scores within 14 days starting at 30/90 days after the first utilization, and WAC, individual score closest to 30/90 days after the first utilization within a range of +/-10 days

Inferential Statistics

Since, as of yet, research findings concerning our research questions are inconclusive, modeling relationships among the several variables in complex ways appeared inadvisable. Therefore, we chose a bivariate approach. When the independent variable was binary, we performed two-tailed Welch's t tests when the dependent variable was metrically scaled and Wilcoxon signed rank/Mann-Whitney U tests when it was ordinally scaled. When both the independent variable and the dependent variable were at least ordinally scaled, we calculated Pearson correlation coefficients or Spearman correlation coefficients, respectively. For each research question, tests were performed on an alpha level of 5%. Holm-corrected P values are reported.

Results

Summary

After a brief description of the baseline characteristics of our sample and the general usage behavior, the results of the analyses concerning our research questions are shown below.

Baseline Characteristics

The statistical analyses were conducted on a final sample of the 306 participants (females, n=170, 55.6%) of the intervention group, who completed the assessments at all 3 time points. The age range of these participants was 18 to 69 years with a mean (SD) age of 40.36 (11.15) years. Of the 306 participants, 128 (41.8%) reported to be in a relationship and 92 (30.1%) reported having at least one child. A total of 119 (38.9%) participants reported achieving a general higher education entrance qualification (Abitur), and 78 (25.5%) finished academic studies. A total of 259 (86.6%) participants were working at the time of the baseline assessment. Further information on the baseline demographics and user characteristics are shown in Table 1.

Table 1. Overview of the baseline characteristics of the final sample.

Measure	Participants (n)	Mean (SD)	Median (range)
Age (years)	306	40.36 (11.15)	40 (18-69)
AFC[a] (years)	305	15.04 (3.23)	15 (2-45)
ARC[b] (years)	305	17.5 (4.21)	17 (0-45)
QFI-days[c]	306	23.05 (8.31)	27 (0-30)
TFB[d]	306	52.91 (56.68)	41.64 (0-515.71)
BDD[e]	304	16.79 (11.45)	20 (0-30)
RR-I[f]	296	8.11 (1.77)	8 (1-10)
RR-C[g]	296	6.08 (2.44)	6 (1-10)
AASE-T[h]	298	62.26 (15.22)	60 (23-93)
AASE-C[i]	296	55.82 (16.1)	59 (20-98)
CAEQ-A[j]	299	7.69 (3.18)	8 (3-15)
CAEQ-CP[k]	299	11.64 (3.1)	12 (4-20)
CAEQ-SE[l]	299	8.19 (3.02)	9 (3-15)
CAEQ-SP[m]	299	16.28 (3.96)	16 (5-25)
CAEQ-T[n]	299	13.21 (2.89)	13 (4-20)
RCQ-C[o]	298	14.54 (3.57)	14 (4-20)
RCQ-P[p]	298	8.7 (3.47)	9 (4-19)
RCQ-A[q]	298	12.2 (3.43)	12 (4-20)

[a]AFC: age of first alcohol consumption.

[b]ARC: age of regular alcohol consumption.

[c]QFI-Days: drinking days measured with the Quantity-Frequency Index.

[d]TFB: average daily alcohol consumption measured with the Timeline Followback approach.

[e]BDD: binge drinking day.

[f]RR-I: readiness ruler importance scale.

[g]RR-C: readiness ruler confidence scale.

[h]AASE-T: Temptation scale of the Alcohol Abstinence Self-Efficacy Scale.

[i]AASE-C: Confidence scale of the Alcohol Abstinence Self-Efficacy Scale.

[j]CAEQ-A: Aggression scale of the Comprehensive Alcohol Expectancy Questionnaire.

[k]CAEQ-CP: Cognitive impairment and physical discomfort scale of the Comprehensive Alcohol Expectancy Questionnaire.

[l]CAEQ-SE: Sexual enhancement scale of the Comprehensive Alcohol Expectancy Questionnaire.

[m]CAEQ-SP: Social assertiveness and positive affect scale of the Comprehensive Alcohol Expectancy Questionnaire.

[n]CAEQ-T: Tension reduction scale of the Comprehensive Alcohol Expectancy Questionnaire.

[o]RCQ-C: Contemplation scale of the Readiness to Change Questionnaire.

[p]RCQ-P: Precontemplation scale of the Readiness to Change Questionnaire.

[q]RCQ-A: Action scale of the Readiness to Change Questionnaire.

Measures of Engagement

While 61 of the total 306 (19.9%) participants did not log in to the website even once, only 6 of the 183 (3.3%) participants who reported their usage at T2 reported not having used the intervention at all. At T1, the most chosen answer (73/200, 36.5%) for the SRU was "weekly 3-4 hours," which would approximately be at least 2160 minutes in a 90-day period. However, based on the FMB, the maximum time spent in the intervention within the first 90 days was 845 minutes. Table 2 shows that, after this period, in the following 90 days (T2) of the 180-day usage period, the usage of the intervention declined for all system usage metrics. Concerning the intervention questionnaires, we found that the MC questionnaire was used less often than the DAC as well as the weekly AUDIT-C. For an overview of the usage statistics, see Table 2.

Table 2. Overview of the engagement with the digital behavior change intervention as captured by all measures available.

Measures and metrics	T1[a]	T2[b]	TOT[c]
Five-minute blocks of activity			
Minimum	0	0	0
Median (1st quartile, 3rd quartile)	205 (30, 275)	20 (0, 25)	240 (30, 295)
Mean (SD)	184.4 (150.3)	25.15 (55.71)	209.5 (184.64)
Maximum	845	460	1105
Participants (N)	306	306	306
Daily alcohol check frequency			
Minimum	0	0	0
Median (1st quartile, 3rd quartile)	22 (1, 23)	3 (0, 3)	25 (1, 26)
Mean (SD)	16.77 (15.89)	3.72 (10.04)	20.49 (23.29)
Maximum	83	90	155
Participants (N)	306	306	306
Weekly alcohol check frequency			
Minimum	0	0	0
Median (1st quartile, 3rd quartile)	12 (1, 13)	1 (0, 1)	13 (1, 14)
Mean (SD)	7.94 (5.81)	1.11 (2.31)	9.05 (7.05)
Maximum	16	25	37
Participants (N)	306	306	306
Mood check frequency			
Minimum	0	0	0
Median (1st quartile, 3rd quartile)	1 (1, 4)	0 (0, 0)	1 (1, 4)
Mean (SD)	5.98 (12.9)	1.12 (5.47)	7.1 (17.05)
Maximum	80	68	147
Participants (N)	306	306	306
Self-reported usage			
Minimum	0	0	N/A[d]
Median (1st quartile, 3rd quartile)	3 (2, 4)	3 (2, 3)	N/A
Mean (SD)	2.82 (1.19)	2.59 (1.12)	N/A
Maximum	5	4	N/A
Participants (N)	200	183	N/A

[a]T1: period from baseline to first post-assessment (90 days).

[b]T2: period from first to second post-assessment (90 days).

[c]TOT: total usage period (180 days).

[d]N/A: not applicable.

Engagement and Outcomes

All SUD except the MC count at T2 and the WAC count at T1 were correlated with one another, with coefficients between r=0.26 and r=0.98. SRU T1 was significantly positively correlated with FMB T1 (ρ=0.29, *P*=.005), FMB TOT (ρ=0.3, *P*=.003), DAC T1 (ρ=0.29, *P*=.007), and WAC T1 (ρ=0.44, *P*<.001). WAC T1 was also positively correlated with SRU T2 (ρ=0.38, *P*<.001). However, the MC was negatively correlated

with SRU T1 and SRU T2 for all 3 intervals with coefficients between ρ=–0.26 and ρ=–0.34.

Significant correlations between changes in drinking behavior and SUD were rare. Participants who utilized the WAC more frequently within the first 90 days (T1) appeared to have achieved higher reductions in BDDs within this period (r=0.26, *P*=.03) as well as in total (r=0.36, *P*<.001). However, more frequent use of the WAC after the first 90 days was associated with smaller total reductions in BDDs (r=–0.28, *P*=.02). A negative relationship between the total reductions in BDDs was

also found for MC T1 (r=–0.39, *P*<.001) and MC TOT (r=–0.37, *P*<.001). Changes in average daily alcohol consumption (TFB) were not significantly correlated with SUD.

In contrast to the SUD, both SRU T1 and SRU T2 were associated with changes in TFB as well as in BDDs. SRU at T1 was positively correlated with BDDs T1 (ρ=0.62, *P*<.001), BDD TOT (ρ=0.65, *P*<.001), TFB T1 (ρ=0.39, *P*<.001), and TFB TOT (ρ=0.42, *P*<.001). SRU T2 was associated with BDD T1 (ρ=0.27, *P*=.03), BDD T2 (ρ=0.3, *P*=.006), BDD TOT (ρ=0.42, *P*<.001), and TFB T2 (ρ=0.29, *P*=.015). This indicates that those participants who reported higher usage also reported higher reductions in BDDs as well as in average daily alcohol consumption. A table including all bivariate correlations can be found in Multimedia Appendix 2.

Baseline User Characteristics and Engagement

We first described the results of the analyses concerning relationships between sociodemographic baseline characteristics and follow-up with the results of the analyses on clinical and psychological baseline characteristics.

Sociodemographic User Characteristics

User age was not correlated with usage. Neither Pearson correlation coefficients for age and FMB at T1 (r=0.16, *P*=.09) and FMB at T2 (r=0.11, *P*=.82) nor the Spearman coefficients for age and SRU at T1 (ρ=–0.13, *P*=.82) and SRU T2 (ρ=–0.2; *P*=.11) were significantly different from zero after Holm-correction of *P* values.

The only significant differences on the system usage measure (FMB) were found between singles and nonsingles (*P*=.02) at T1 as well as between those who did not complete secondary education and those who did (*P*=.02) at T1 (Table 3).

However, for the subjective measure, for every demographic, except gender, the groups differed significantly (Table 3). Higher usage was reported from singles (vs nonsingles) (T1: W=6883.5, *P*<.001; T2: W=4832, *P*<.001), users without children (vs with one child or more) (T1: W=5003, *P*<.001; T2: W=3801.5, *P*<.001), users who did not start or finish secondary education (vs with secondary education) (T1: W=6596.5, *P*<.001; T2: W=5300, *P*<.001), users without academic education (vs academic education) (T1: W=4622.5, *P*<.001, W=3656, *P*<.001) and those who worked (vs no work) (T1: W=1272.5, *P*=.001; T2: W=774, *P*=.004) at T1 and T2. These results must be interpreted cautiously because group sizes tended to differ substantially.

Table 3. Relationships between sociodemographic user characteristics and measures of engagement.

Measures, Groups	Mean (SD)	t (df)[a]	W[b]	P value[c]
FMB[d] T1[e], Gender		0.44 (295.52)	N/A[f]	>.99
Females (n=170)	187.74 (154.11)			
Males (n=136)	180.18 (145.86)			
FMB T2[g], Gender		–0.13 (291.19)	N/A	>.99
Females (n=170)	24.76 (56.17)			
Males (n=136)	25.63 (55.35)			
SRU[h] T1, Gender		N/A	5075.5	>.99
Females (n=113)	2.78 (1.24)			
Males (n=87)	2.87 (1.12)			
SRU T2, Gender		N/A	4563	>.99
Females (n=102)	2.5 (1.12)			
Males (n=81)	2.7 (1.1)			
FMB T1, Relationship status		3.31 (198.94)	N/A	.02
Single (n=178)	209.80 (115.76)			
Nonsingle (n=128)	149.02 (182.79)			
FMB T2, Relationship status		–0.37 (154.48)	N/A	>.99
Single (n=178)	24.02 (30.36)			
Nonsingle (n=128)	26.72 (78.52)			
SRU T1, Relationship status		N/A	6883.5	<.001
Single (n=142)	3.24 (0.99)			
Nonsingle (n=58)	1.79 (1.00)			
SRU T2, Relationship status		N/A	4832	<.001
Single (n=140)	2.89 (.93)			
Nonsingle (n=43)	1.61 (1.09)			
FMB T1, With/Without children		1.39 (128.68)	N/A	>.99
No children (n=214)	193.41 (129.18)			
One or more children (n=92)	163.37 (189.78)			
FMB T2, With/Without children		–0.07 (127.66)	N/A	>.99
No children (n=214)	24.98 (47.78)			
One or more children (n=92)	25.54 (71.14)			
SRU T1, With/Without children		N/A	5003	<.001
No children (n=159)	3.05 (1.12)			
One or more children (n=41)	1.93 (1.01)			
SRU T2, With/Without children		N/A	3801.5	<.001
No children (n=151)	2.8 (1.03)			
One or more children (n=32)	1.63 (0.98)			
FMB T1, Educational qualification		–3.38 (200.88)	N/A	.02
No Abitur (n=187)	208.57 (129.06)			
Abitur (n=119)	146.39 (172.49)			
FMB T2, Educational qualification		–0.004 (176.6)	N/A	>.99
No Abitur (n=187)	25.16 (43.94)			

Measures, Groups	Mean (SD)	t (df)[a]	W[b]	P value[c]
Abitur (n=119)	25.13 (70.57)			
SRU T1, Educational qualification		N/A	6596.5	<.001
No Abitur (n=145)	3.21 (.99)			
Abitur (n=55)	1.78 (1.03)			
SRU T2, Educational qualification		N/A	5300	<.001
No Abitur (n=138)	2.95 (.92)			
Abitur (n=45)	1.49 (.92)			
FMB T1, Job qualification		–2.93 (115.15)	N/A	.07
Nonacademic field (n=228)	200.26 (140.12)			
Academic field (n=78)	137.95 (169.33)			
FMB T2, Job qualification		0.48 (93.81)	N/A	>.99
Nonacademic field (n=228)	23.99 (44.58)			
Academic field (n=78)	28.53 (80.13)			
SRU T1, Job qualification		N/A	4622.5	<.001
Nonacademic field (n=166)	3.06 (1.08)			
Academic field (n=34)	1.65 (.98)			
SRU T2, Job qualification		N/A	3656	<.001
Nonacademic field (n=153)	2.79 (1.02)			
Academic field (n=30)	1.57 (1.01)			
FMB T1, Job status		–0.87 (55.48)	N/A	>.99
No work (n=47)	206.06 (191.52)			
Work (n=259)	180.44 (141.65)			
FMB T2, Job status		–1.09 (52.93)	N/A	>.99
No work (n=47)	36.17 (79.13)			
Work (n=259)	23.15 (50.27)			
SRU T1, Job status		N/A	1272.5	.001
No work (n=28)	2 (.94)			
Work (n=172)	2.95 (1.17)			
SRU T2, Job status		N/A	774	.004
No work (n=19)	1.68 (1)			
Work (n=164)	2.7 (1.08)			

[a]Welch's t test.

[b]Wilcoxon signed rank/Mann-Whitney U tests.

[c]P values are Holm-corrected.

[d]FMB: five-minute blocks of activity.

[e]T1: period from baseline to first post-assessment (90 days).

[f]N/A: not applicable.

[g]T2: period from first to second post-assessment (90 days).

[h]SRU: self-reported usage.

Clinical and Psychological User Characteristics

Concerning SUD, only more drinking days (r=0.29, P<.001) and more BDDs (r=0.27, P<.001) were associated with more time spent using the intervention within the first 90 days (T1). However, for the self-report measure, multiple correlation coefficients were significantly different from zero. Users with a higher number of drinking days reported higher usage within the first (T1: ρ=0.37, P<.001) and second (T2: ρ=0.6, P<.001) 90 days of access to the intervention. The same was true with regard to the number of BDDs (T1: ρ=0.52, P<.001; T2:ρ=0.56, P<.001). Those with an older age when they first consumed

alcohol reported higher usage at both T1 (ρ=0.34, P<.001) and T2 (ρ=0.43, P<.001).

Higher scores on the temptation scale of the Alcohol Abstinence Self-Efficacy Scale were significantly positively correlated with higher reported usage at T2 (ρ=0.46, P<.001) but not at T1 (ρ=0.19, P>.99). From the comprehensive alcohol expectancy questionnaire, the aggression scale was significantly correlated with SRU at T1 (ρ=0.34, P<.001) and T2 (ρ=0.47, P<.001). The same was true for the sexual enhancement scale of the comprehensive alcohol expectancy questionnaire (T1: ρ=0.38, P<.001; T2: ρ=0.54, P<.001). Higher scores on the confidence scale of the readiness ruler were associated with higher reported usage at T1 (ρ=0.26, P=.049). While lower scores on the contemplation (T1: ρ=–0.43, P<.001; T2: ρ=–0.31, P=.008)

and higher scores on the precontemplation scale (T1: ρ=0.57, P<.001; T2: ρ=0.43, P<.001) of the readiness to change questionnaire were associated with higher SRU at T1 and T2, no significant association was found for scores on the action scale (T1: ρ=–0.07, P>.99; T2: ρ=0.09, P>.99). All correlations that were calculated can be found in Multimedia Appendix 3.

Intervention Questionnaires and Outcomes

For the 30-day period, only changes in the WAC scores were correlated with total reductions in the TFB (r=0.24, P=.015). However, for the 90-day period, changes in WAC scores were correlated with changes in TFB at T1 (r=0.31, P<.001) as well as with total changes (r=0.39, P<.001) and changes in the DAC were positively correlated with all 4 outcome measures with coefficients between r=0.34 and r=0.58 (Table 4).

Table 4. Correlations between intervention questionnaires and assessment questionnaires.

Measures	WAC[a] 30D[b]	WAC 90D	DAC[c] 30D	DAC 90D	TFB[d] T1[e]	TFB TOT[f]	BDD[g] T1	BDD TOT
WAC 30D								
n[h]	200	168	189	170	200	200	175	169
r	1	0.69	0.31	0.16	0.14	0.24	0.01	−0.04
P value	<.001	<.001	<.001	.66	.73	.02	>.99	>.99
WAC 90D								
n	168	169	165	165	165	169	163	159
r	0.69	1	0.35	0.34	0.31	0.39	0.23	0.18
P value	<.001	<.001	<.001	<.001	.001	<.001	.06	.48
DAC 30D								
n	189	165	191	169	191	191	170	164
r	0.31	0.35	1	0.24	0.04	−0.02	0.08	0.01
P value	<.001	<.001	<.001	.03	>.99	>.99	>.99	>.99
DAC 90D								
n	170	165	169	173	173	173	165	161
r	0.16	0.34	0.24[i]	1	0.34	0.36	0.58	0.51
P value	.66	<.001	.03	<.001	<.001	<.001	<.001	<.001
TFB T1								
n	200	165	191	173	306	306	204	189
r	0.14	0.31	0.04	0.34	1	0.84	0.42	0.37
P value	.73	.001	>.99	<.001	<.001	<.001	<.001	<.001
TFB TOT								
n	200	169	191	173	306	306	204	187
r	0.24	0.39	−0.02	0.36	0.84	1	0.33	0.58
P value	.02	<.001	>.99	<.001	<.001	<.001	<.001	<.001
BDD T1								
n	175	163	170	165	204	204	204	189
r	0.01	0.23	0.08	0.58	0.42	0.33	1	0.68
P value	>.99	.06	>.99	<.001	<.001	<.001	<.001	<.001
BDD TOT								
n	169	159	164	161	189	187	189	189
r	−0.04	0.18	0.01	0.51	0.37	0.58	0.68	1
P value	>.99	.48	>.99	<.001	<.001	<.001	<.001	<.001

[a]WAC: weekly alcohol check.

[b]D: time span from baseline in days.

[c]DAC: daily alcohol check.

[d]TFB: average daily alcohol consumption measured with the Timeline Followback approach.

[e]T1: period from baseline to first post-assessment (90 days).

[f]TOT: total usage period (180 days).

[g]BDD: binge drinking day.

[h]Associated number of cases.

[i]The correlation was significant at a significance level of .05 (two-tailed).

XSL•FO
RenderX

Discussion

Principal Findings

To analyze engagement with Vorvida, we explored the following 3 questions.

Engagement and outcomes: How is engagement with the intervention related to drinking outcomes?

User characteristics and engagement: Which user demographics and clinical or psychological baseline characteristics are associated with measures of engagement?

Intervention questionnaires and outcomes: Are reported outcomes associated with data captured by voluntary intervention questionnaires?

We used SUD as well as SRU to measure engagement. The amount of time spent engaging with the intervention as measured by SUD did not match the SRU. Users tended to overestimate their actual usage of the intervention. This is in line with the general observation that self-reports of internet and smartphone use tend to be inaccurate [29,30]. More specifically, Perski et al [31] showed that the amount of usage of a smartphone app for reducing alcohol consumption as measured by self-report did not correlate with server log data. However, in our study, at least for the first 90 days of intervention usage, both measures correlated positively with one another. After the first 90 days, average engagement with the intervention declined on all measures used.

Engagement and Outcomes

Similar to that reported in other studies [8,10,11], no conclusive relationships could be established between engagement as measured by SUD and changes in drinking behavior. While the time spent using the intervention as measured by SUD was not associated with changes in drinking behavior, the results concerning relationships between the frequency of the utilization of the intervention questionnaires and changes in drinking behavior remained inconclusive. However, SRU was correlated positively with reductions in average daily alcohol consumption as well as reductions in BDDs.

An explanation for this phenomenon, following Yardley et al [1], would be that self-reports capture aspects of engagement with behavior change, apart from the actual usage of the website. When asked to rate how often they used Vorvida, apart from the actual time spent on the website, time spent engaging with printed homework material and in applying newly learned techniques to real-life situations might have influenced the participants' choices. It seems reasonable to assume that those participants who spent more time and effort actually implementing the newly learned skills would report higher drinking reductions as well as higher usage. Since the SUD does not capture time spent engaging with behavior change, this might also explain why SRU was overall higher than usage captured by the SUD.

User Characteristics

Regarding user demographics, except age and gender, all other measures were associated with SRU. Higher usage was reported from singles, users without children, users without secondary education, as well as users without academic education, and users who work. With regard to the time spent using the intervention as captured by server logs, significantly higher usage was recorded only for singles and users without secondary education.

The facts that in our analyses, neither age nor gender were associated with higher engagement and participants with higher education engaged less with the intervention than those with lower education levels stand in contrast to previous findings [11,16,17]. In addition to this, we found that singles and users without children tended to engage more with the intervention, though only tests for differences in the self-report measures were significant. Nonetheless, one could argue that these associations are due to singles and persons without children having more free time to use the intervention and higher educated users needing less time to understand and implement the intervention content.

Concerning clinical and psychological baseline characteristics, we also found multiple associations with SRU, while only the number of drinking days and BDDs were positively correlated with higher SUD during the first usage period. In addition to more drinking days and more BDDs, first consumption at an older age was associated with higher SRU. Since we found more drinking days and more BDDs at baseline to be positively related to engagement but no relationship for average daily alcohol consumption, our results on baseline problem severity add to the mixed findings of previous studies [11,17].

Participants reporting more temptation to drink alcohol in different situations also reported higher usage at T2 but not at T1. However, confidence to withstand drinking in these situations was not associated with reports of usage. Unlike the findings of Murray et al [16], we found higher confidence in one's ability to change to be associated with higher engagement. Participants who were more confident in their ability to achieve a reduction reported higher engagement within the first usage period. With respect to alcohol expectations, only those related to aggression or sexual enhancement were correlated with engagement. Interestingly, higher scores on the precontemplation scale and lower scores on the contemplation scale were associated with higher SRU. This suggests that those participants who were already thinking about reducing their drinking, used the intervention less than those who had no intention to change. Since individuals in the precontemplation stage are not intending to change their drinking yet [29], one would expect them to engage less with the intervention than those already contemplating about changing. Consistent with this assumption, previous studies found readiness to change to be associated with higher engagement [17]. However, our results point in a different direction. While higher precontemplation scores were associated with higher engagement, the opposite was true for contemplation scores.

Since all clinical and psychological measures related to engagement (except confidence to achieve behavior change) were correlated with one another, an underlying factor might be responsible for their relationship with SRU. A potential factor that comes to mind is impulsivity. Personality traits related to

impulsivity have been robustly linked to alcohol use, especially drinking frequency, binge drinking, and alcohol-related problem behavior [32-34]. Furthermore, aggressive behavior [32-34] and sexual risk taking [3-41] have been linked to aspects of impulsiveness and alcohol-induced intoxication. This might explain why only the aggression-related and sexual enhancement-related alcohol outcome expectancies were found to be associated with higher reports of engagement. Concerning readiness to change, higher ratings of impulsivity have been linked to lower stages of change [33]. Assuming that users with high levels of impulsivity engaged more often with the intervention than those with low levels, it still does not explain why they did so. On the one hand, an impulsive person might have trouble implementing the techniques taught by Vorvida and therefore might continue using the intervention to get further help. On the other hand, higher effect sizes for the number of BDDs than the average daily alcohol consumption found in the RCT [20] suggest that the intervention content might be more suitable for reducing excessive drinking than regular drinking. Therefore, impulsive users might have engaged more with the intervention due to a better fit for their needs.

Intervention Questionnaires and Outcomes

Changes in alcohol consumption as reported in the voluntary intervention questionnaires were found to be associated with changes captured by our assessments. While larger reductions captured by the weekly questionnaire within the first 30 days were correlated only with higher reductions in average daily consumption after 180 days, higher reductions in WAC scores within the first 30 days were associated with higher reductions in daily consumption after 90 as well as 180 days. However, higher reductions captured by the daily questionnaire within the first 90 days were associated with higher reductions in average daily consumption and in the number of BDDs after 90 and 180 days.

Therefore, we could show that individual changes in drinking behavior as captured by the intervention questionnaires were associated with changes captured by our assessments. Though our analyses are only a first step toward establishing a connection between intervention questionnaires and actual behavior change, they support the idea that intervention questionnaires could potentially be used to aid dynamic tailoring. Based on data provided by the users concerning their progress, those who struggle with reducing their drinking could receive further motivational content while for those who already reduced their drinking, relapse prevention content could be provided.

Strengths and Limitations

Our analyses are based on data that were collected as part of an RCT. Since the RCT was not designed to primarily answer questions of engagement, we have to report several limitations. SUD only captured the total time a user spent in the intervention and how often they used the questionnaires. Thus, we could not capture whether the individual's usage was distributed over few or many sessions or which modules were completed. The fact that SUD did rarely correlate with any of the outcomes or user characteristics might therefore be due to different patterns of use resulting in the same amount of time spent in the intervention [7,12]. Because the questionnaire that we used to

measure SRU has not been evaluated, its validity remains uncertain. At the time the questionnaire was developed, many of the central works that advanced the theoretical understanding of engagement had not yet been published, eg, the work of Yeager and Benight [14]. Therefore, some important concepts of engagement could not be considered in the context of operationalization of engagement. Furthermore, since participants were asked to report their intervention usage over the previous 3 months, answers might be influenced by recall bias [1], which could limit the validity of the results. This limitation of the validity of the results could be an alternate explanation for the discrepancy between subjective and objective usage. This limitation of the validity of the results could be an alternate explanation for the discrepancy between subjective and objective usage. In addition to this, as social desirability bias is a potential problem for all self-report measures [42], our results might be influenced by social desirability. Nonetheless, our study is one of the few studies to analyze relationships between engagement and changes in drinking behavior utilizing SRU as well as SUD to measure engagement. Furthermore, we provide analyses of relationships between engagement and a wide range of different demographic, clinical, and psychological variables based on a fairly large sample.

Conclusions

Overall, our results concerning engagement-outcome relationships and associations between user characteristics and engagement add to the inconclusive evidence that can be found across the literature [1,8,14]. SRU, unlike SUD, was consistently related to better outcomes. This suggests that our questionnaire also captured other dimensions of engagement apart from the time spent using the intervention and that these dimensions are crucial for behavior change. On top of this, we found SRU to be associated with multiple, intercorrelated user characteristics, most of which have been linked to impulsivity in the past. This indicates that Vorvida might be especially engaging to impulsive persons. However, since DBCIs are made available for a wide range of target behaviors, with different content and features, it is uncertain whether this result can be generalized to other DBCIs. Finally, we could show that changes in the target behavior as captured by intervention questionnaires were related to changes reported in our assessments.

Recommendations for Future Research

As evidence for the efficacy of DBCIs grows, it becomes more important to explore the question of how DBCIs bring about behavior change. A key part of this endeavor will be to gain a better understanding of the various facets of engagement and how they relate to behavior change. Different frameworks and theoretical models have recently been proposed to guide engagement-based research [1,6,14]. To empirically test these models, it will be crucial to measure the multidimensional nature of engagement in a reliable and valid way. Therefore, psychometrically sound self-report questionnaires that capture aspects of the micro and macro level as well as qualitative aspects of engagement need to be developed and tested in future studies. Since it might be difficult to remember usage behavior over an extended period, we encourage timely assessment of the intensity of use. This would allow measurement of

engagement and early responses to the intervention in the moment, avoid recall bias, and could potentially advance tailoring. To identify universal predictors of engagement, it appears reasonable to shift the focus from demographic and problem behavior–specific variables to theory-based constructs as Yeager and Benight have suggested [14].

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of all measures used in the analyses.
[PDF File (Adobe PDF File), 48 KB - jmir_v22i11e18826_app1.pdf]

Multimedia Appendix 2
Correlations between measures of engagement and changes in drinking behavior.
[PDF File (Adobe PDF File), 34 KB - jmir_v22i11e18826_app2.pdf]

Multimedia Appendix 3
Correlations between clinical as well as psychological baseline characteristics and measures of engagement.
[PDF File (Adobe PDF File), 36 KB - jmir_v22i11e18826_app3.pdf]

References

1. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
2. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
3. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
4. Cavanagh K. Turn on, tune in and (don't) drop out: Engagement, adherence, attrition, and alliance with internet-based interventions. In: Bennett-Levy J, Richards DA, Farrand P, Christensen H, Griffiths KM, Kavanaugh DJ, et al, editors. Oxford guide to low intensity CBT interventions. Oxford: Oxford University Press; 2010:227-233.
5. Ryan C, Bergin M, Wells JS. Theoretical Perspectives of Adherence to Web-Based Interventions: a Scoping Review. Int J Behav Med 2018 Feb;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
6. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
7. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
8. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
9. Johansson M, Sinadinovic K, Hammarberg A, Sundström C, Hermansson U, Andreasson S, et al. Web-Based Self-Help for Problematic Alcohol Use: a Large Naturalistic Study. Int J Behav Med 2017 Oct;24(5):749-759 [FREE Full text] [doi: 10.1007/s12529-016-9618-z] [Medline: 27900733]
10. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
11. Garnett C, Perski O, Tombor I, West R, Michie S, Brown J. Predictors of Engagement, Response to Follow Up, and Extent of Alcohol Reduction in Users of a Smartphone App (Drink Less): Secondary Analysis of a Factorial Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11175 [FREE Full text] [doi: 10.2196/11175] [Medline: 30552081]

12. Milward J, Deluca P, Drummond C, Kimergård A. Developing Typologies of User Engagement With the BRANCH Alcohol-Harm Reduction Smartphone App: Qualitative Study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e11692 [FREE Full text] [doi: 10.2196/11692] [Medline: 30545806]

13. Lenhard F, Mitsell K, Jolstedt M, Vigerland S, Wahlund T, Nord M, et al. The Internet Intervention Patient Adherence Scale for Guided Internet-Delivered Behavioral Interventions: Development and Psychometric Evaluation. J Med Internet Res 2019 Oct 01;21(10):e13602 [FREE Full text] [doi: 10.2196/13602] [Medline: 31573901]

14. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]

15. Beatty L, Binnion C. A Systematic Review of Predictors of, and Reasons for, Adherence to Online Psychological Interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]

16. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]

17. Postel MG, de Haan HA, ter Huurne ED, van der Palen J, Becker ES, de Jong CA. Attrition in web-based treatment for problem drinkers. J Med Internet Res 2011 Dec 27;13(4):e117 [FREE Full text] [doi: 10.2196/jmir.1811] [Medline: 22201703]

18. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]

19. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: A mixed-methods study. Digit Health 2018;4:2055207618785841. [doi: 10.1177/2055207618785841] [Medline: 31463077]

20. Zill J, Christalle E, Meyer B, Härter M, Dirmaier J. The Effectiveness of an Internet Intervention Aimed at Reducing Alcohol Consumption in Adults. Dtsch Arztebl Int 2019 Feb 22;116(8):127-133 [FREE Full text] [doi: 10.3238/arztebl.2019.0127] [Medline: 30940341]

21. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]

22. Zill JM, Meyer B, Topp J, Daubmann A, Härter M, Dirmaier J. Vorvida: study protocol of a randomized controlled trial testing the effectiveness of Internet-based self-help program for the reduction of alcohol consumption for adults. BMC Psychiatry 2016 Jan 29;16:19 [FREE Full text] [doi: 10.1186/s12888-016-0725-9] [Medline: 26822579]

23. Bloomfield K, Hope A, Kraus L. Alcohol survey measures for Europe: A literature review. Drugs: Education, Prevention and Policy 2012 Jan 09;20(5):348-360. [doi: 10.3109/09687637.2011.642906]

24. Demmel R. Motivational Interviewing. In: Linden M, Hautzinger M, editors. Verhaltenstherapiemanual. Berlin: Springer-Verlag; 2015:193-197.

25. DiClemente CC, Carbonari JP, Montgomery RP, Hughes SO. The Alcohol Abstinence Self-Efficacy scale. J Stud Alcohol 1994 Mar;55(2):141-148. [doi: 10.15288/jsa.1994.55.141] [Medline: 8189734]

26. Nicolai J, Demmel R, Moshagen M. The comprehensive alcohol expectancy questionnaire: confirmatory factor analysis, scale refinement, and further validation. J Pers Assess 2010 Sep;92(5):400-409. [doi: 10.1080/00223891.2010.497396] [Medline: 20706926]

27. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Br J Addict 1992 May;87(5):743-754. [doi: 10.1111/j.1360-0443.1992.tb02720.x] [Medline: 1591525]

28. Berner MM, Kriston L, Bentele M, Härter M. The alcohol use disorders identification test for detecting at-risk drinking: a systematic review and meta-analysis. J Stud Alcohol Drugs 2007 May;68(3):461-473. [doi: 10.15288/jsad.2007.68.461] [Medline: 17446987]

29. Scharkow M. The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data. Communication Methods and Measures 2016 Mar 24;10(1):13-27. [doi: 10.1080/19312458.2015.1118446]

30. Boase J, Ling R. Measuring Mobile Phone Use: Self-Report Versus Log Data. J Comput-Mediat Comm 2013 Jun 10;18(4):508-519. [doi: 10.1111/jcc4.12021]

31. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study. J Med Internet Res 2019 Nov 20;21(11):e16197 [FREE Full text] [doi: 10.2196/16197] [Medline: 31746771]

32. Coskunpinar A, Dir AL, Cyders MA. Multidimensionality in impulsivity and alcohol use: a meta-analysis using the UPPS model of impulsivity. Alcohol Clin Exp Res 2013 Sep;37(9):1441-1450 [FREE Full text] [doi: 10.1111/acer.12131] [Medline: 23578176]

33. Kazemi DM, Wagenfeld M, Van Horn RK, Levine MJ, Dmochowski J. Binge Drinking Among Underage College Students: Role of Impulsivity and the Transtheoretical Model. Journal of Addictions Nursing 2011;22(4):193-199. [doi: 10.3109/10884602.2011.616605]

XSL•FO

RenderX

34. Littlefield AK, Stevens AK, Sher KJ. Impulsivity and Alcohol Involvement: Multiple, Distinct Constructs and Processes. Curr Addict Rep 2014 Mar;1(1):33-40 [FREE Full text] [doi: 10.1007/s40429-013-0004-5] [Medline: 29963362]

35. Prochaska JO, Redding CA, Evers KE. The transtheoretical model and stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. San Francisco: Jossey-Bass; 2008:97-121.

36. Crane CA, Godleski SA, Przybyla SM, Schlauch RC, Testa M. The Proximal Effects of Acute Alcohol Consumption on Male-to-Female Aggression: A Meta-Analytic Review of the Experimental Literature. Trauma Violence Abuse 2016 Dec;17(5):520-531. [doi: 10.1177/1524838015584374] [Medline: 26009568]

37. Duke AA, Smith KMZ, Oberleitner LMS, Westphal A, McKee SA. Alcohol, drugs, and violence: A meta-meta-analysis. Psychology of Violence 2018 Mar;8(2):238-249. [doi: 10.1037/vio0000106]

38. Lejuez C, Magidson J, Mitchell S, Sinha R, Stevens M, de Wit H. Behavioral and biological indicators of impulsivity in the development of alcohol use, problems, and disorders. Alcohol Clin Exp Res 2010 Aug;34(8):1334-1345 [FREE Full text] [doi: 10.1111/j.1530-0277.2010.01217.x] [Medline: 20491733]

39. Charnigo R, Noar SM, Garnett C, Crosby R, Palmgreen P, Zimmerman RS. Sensation seeking and impulsivity: combined associations with risky sexual behavior in a large sample of young adults. J Sex Res 2013;50(5):480-488 [FREE Full text] [doi: 10.1080/00224499.2011.652264] [Medline: 22456443]

40. Lan C, Scott-Sheldon LAJ, Carey KB, Johnson BT, Carey MP. Prevalence of Alcohol Use, Sexual Risk Behavior, and HIV Among Russians in High-Risk Settings: a Systematic Review and Meta-Analysis. Int J Behav Med 2017 Apr;24(2):180-190 [FREE Full text] [doi: 10.1007/s12529-016-9596-1] [Medline: 27730501]

41. Scott-Sheldon LAJ, Carey KB, Cunningham K, Johnson BT, Carey MP, MASH Research Team. Alcohol Use Predicts Sexual Decision-Making: A Systematic Review and Meta-Analysis of the Experimental Literature. AIDS Behav 2016 Jan;20 Suppl 1:S19-S39 [FREE Full text] [doi: 10.1007/s10461-015-1108-9] [Medline: 26080689]

42. van de Mortel TF. Faking It: Social Desirability Response Bias in Self-report Research. The Australian Journal of Advanced Nursing. 2008. URL: https://research-repository.griffith.edu.au/handle/10072/56416 [accessed 2020-01-08]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Concise
BDD: binge drinking day
DAC: daily alcohol check
DBCI: digital behavior change intervention
FMB: five-minute blocks of activity
MC: mood check
RCT: randomized controlled trial
SRU: self-reported usage
SUD: system usage data
TFB: Timeline Followback
WAC: weekly alcohol check

XSL•FO
RenderX

Original Paper

Vaping-Related Mobile Apps Available in the Google Play Store After the Apple Ban: Content Review

Meredith C Meacham[1], MPH, PhD; Erin A Vogel[2], PhD; Johannes Thrul[3], PhD

[1]Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, CA, United States

[2]Stanford Prevention Research Center, Department of Medicine, Stanford University, Stanford, CA, United States

[3]Department of Mental Health, Johns Hopkins School of Public Health, Baltimore, MD, United States

Corresponding Author:
Meredith C Meacham, MPH, PhD
Department of Psychiatry and Behavioral Sciences
University of California San Francisco
401 Parnassus Ave
San Francisco, CA, 94143
United States
Phone: 1 415 206 4253
Email: Meredith.Meacham@ucsf.edu

Abstract

Background: In response to health concerns about vaping devices (eg, youth nicotine use, lung injury), Apple removed 181 previously approved vaping-related apps from the App Store in November 2019. This policy change may lessen youth exposure to content that glamorizes vaping; however, it may also block important sources of information and vaping device control for adults seeking to use vaping devices safely.

Objective: Understanding the types of nicotine and cannabis vaping–related apps still available in the competing Google Play Store can shed light on how digital apps may reflect information available to consumers.

Methods: In December 2019, we searched the Google Play Store for vaping-related apps using the keywords "vape" and "vaping" and reviewed the first 100 apps presented in the results. We reviewed app titles, descriptions, screenshots, and metadata to categorize the intended substance (nicotine or cannabis/tetrahydrocannabinol) and the app's purpose. The most installed apps in each purpose category were downloaded and evaluated for quality and usability with the Mobile App Rating Scale.

Results: Of the first 100 apps, 79 were related to vaping. Of these 79 apps, 43 (54%) were specific to nicotine, 3 (4%) were specific to cannabis, 1 (1%) was intended for either, and for the remaining 31 (39%), the intended substance was unclear. The most common purposes of the apps were making do-it-yourself e-liquids (28/79, 35%) or coils (25/79, 32%), games/entertainment (19/79, 24%), social networking (16/79, 20%), and shopping for vaping products (15/79, 19%). Of the 79 apps, at least 4 apps (5%) paired with vaping devices to control temperature or dose settings, 8 apps (10%) claimed to help people quit smoking using vaping, and 2 apps (3%) had the goal of helping people quit vaping.

Conclusions: The majority of vaping-related apps in the Google Play Store had features either to help users continue vaping, such as information for modifying devices, or to maintain interest in vaping. Few apps were for controlling device settings or assisting with quitting smoking or vaping. Assuming that these Google Play Store apps were similar in content to the Apple App Store apps that were removed, it appears that Apple's ban would have a minimal effect on people who vape with the intention of quitting smoking or who are seeking information about safer vaping via mobile apps.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

The emergence of vaping, or e-cigarette technology, whereby a liquid solution is heated in a portable device until aerosolized and then inhaled, has been greeted with both promise and alarm [1-4]. In response to concerns about lung injury associated with vaping [5] and the increasing prevalence of youth vaping [6,7], Apple removed previously-approved vaping-related apps from its App Store (for iOS) in November 2019 [8], but the competing Google Play Store (for Android) did not remove vaping-related apps. Vaping-related apps may negatively influence youth by glamorizing vaping; however, some apps may benefit adults who use tobacco by providing cessation assistance or harm reduction information. The net balance of benefits and drawbacks from vaping-related apps likely depends on app content.

Vaping Nicotine

Vaping nicotine may be less harmful than smoking combustible cigarettes [9] and may be a potential tool in helping people quit smoking [4,10,11]. Yet, alongside this potential is limited scientific evidence about the efficacy of nicotine vaping in promoting and maintaining tobacco-smoking cessation, as well as mounting concerns about a youth vaping epidemic [12,13]. Evidence to date appears to be mixed and context-dependent regarding whether e-cigarettes help adults quit [10] with concerns that the amount of nicotine in popular vaping products is too high and that people who quit smoking by vaping may return to smoking [1].

E-cigarettes are currently the most popular tobacco product among US youth, with an estimated 1 in 5 US high-school seniors reporting having vaped nicotine in the previous month, and 1 in 4 reporting having vaped nicotine- or flavor-containing e-liquids in 2018 [6]. A higher past-month prevalence of vaping was reported in the 2019 National Youth Tobacco Survey, with more than 1 in 4 US high-school students using e-cigarettes [7]. Although adolescents are smoking fewer cigarettes per day [14], and smoking has become less prevalent [15] in recent years, there are also reports of greater nicotine dependence among adolescents and young adults who vape prefilled pods (cartridges) [16], and a body of research indicating harm to the developing brain from nicotine [17]. In response, national and local governments have taken steps to propose or enact restrictions (or bans) on e-cigarette sales [18,19] and flavored vaping products [20].

Vaping Cannabis

Similarly, vaping cannabis extracts or flower is believed to be less harmful than inhaling combusted cannabis [21-23]. Vaping is an increasingly popular method of cannabis consumption in the United States [24-26] although cannabis remains illegal in many US states and most countries. Both formal and informal cannabis product markets now offer a wide array of ways to consume tetrahydrocannabinol (THC), including vaping products [2,27]. The legalization of cannabidiol from hemp products in the United States in 2018 has led to an increase in interest and availability of cannabidiol-containing products [28] some of which can be vaped and do not need to be legally purchased from a licensed dispensary. Many vaping products with both THC and cannabidiol are sold online with limited regulation, and published product testing studies have found that tested cannabinoid content was often higher or lower than labeled content [29,30]. While Canada legalized cannabis nationally in October 2018, extracts and edibles were not legal until October 2019, due to additional difficulties in regulation.

Against this background, in the spring and summer of 2019, cases of acute lung injury related to vaping began to be reported in the United States. This new illness, named electronic or vaping product use associated lung injury (EVALI), ultimately caused at least 52 deaths and over 2600 hospitalizations by December 2019 [5]. Most of these patients were young adults who were previously healthy and reported a gradual onset of respiratory, constitutional, and gastrointestinal symptoms. Most—though not all—patients reported having used e-cigarette or vaping products containing THC. The most likely culprit was identified as vitamin E acetate, used as a bulking agent in primarily THC-containing vaping products, often procured in a state where cannabis was still illegal or from an unlicensed seller [31-34]. People who vape cannabis or nicotine may turn to mobile apps for guidance on safe use.

Google Play Store Versus Apple iOS App Store

Mobile phones and mobile apps are increasingly used for accessing health and safety information about many topics [35,36], including substances. Apps are most used by younger populations with higher income and education [37]. The Google Play Store for Android and the Apple App Store for iOS are the 2 major mobile app platforms and marketplaces for digital apps, with 2.6 million available apps on Google Play and 1.8 million apps available on the Apple App Store in early 2020 [38]. Many popular apps have versions for both platforms. Both app stores have content, technical, and stylistic guidelines that developers must follow in order to have their app approved; violations can be grounds for removal of a previously approved app. It is generally believed that getting approval from the Apple App Store is more difficult than getting approval from the Google Play Store [39], with Apple App Store apps often viewed as higher quality and less likely to be free to use [40].

Both app stores specifically address tobacco and cannabis content in their developer guidelines (Table 1). Google Play Store's substance-related content guidelines prohibit apps "facilitating the sale" of tobacco, marijuana, alcohol, or illegal drugs, or those "depicting or encouraging" use by minors [41]. The Apple App Store guidelines prohibit apps that "encourage consumption of tobacco and vape products, illegal drugs, or excessive amounts of alcohol," particularly those encouraging minors [42]. Facilitating sale of these substances is also not allowed [42].

XSL•FO

RenderX

Table 1. App Store Review Guidelines related to tobacco, vaping, and marijuana.

App store and section	Text quotation
Apple App Store for iOS	
Safety	*1.4.3 Apps that encourage consumption of tobacco and vape products, illegal drugs, or excessive amounts of alcohol are not permitted on the App Store. Apps that encourage minors to consume any of these substances will be rejected. Facilitating the sale of marijuana, tobacco, or controlled substances (except for licensed pharmacies) isn't allowed.* [42]
Google Play Store for Android	
Illegal Activities	*We don't allow apps that facilitate or promote illegal activities.*
	Examples of common violations
	Facilitating the sale or purchase of illegal drugs or prescription drugs without a prescription.
	Depicting or encouraging the use or sale of drugs, alcohol, or tobacco by minors.
	Instructions for growing or manufacturing illegal drugs. [41]
Inappropriate Content—Marijuana	*We don't allow apps that facilitate the sale of marijuana or marijuana products, regardless of legality.*
	Allowing users to order marijuana through an in-app shopping cart feature.
	Assisting users in arranging delivery or pick up of marijuana.
	Facilitating the sale of products containing THC (Tetrahydrocannabinol), including products such as CBD oils containing THC. [41]
Inappropriate Content—Tobacco & Alcohol	*We don't allow apps that facilitate the sale of tobacco (including e-cigarettes and vape pens) or encourage the illegal or inappropriate use of alcohol or tobacco.*
	Depicting or encouraging the use or sale of alcohol or tobacco to minors.
	Implying that consuming tobacco can improve social, sexual, professional, intellectual, or athletic standing.
	Portraying excessive drinking favorably, including the favorable portrayal of excessive, binge or competition drinking. [41]

Apple iOS App Store's Vaping App Ban

In response to health concerns about vaping devices (eg, youth nicotine use, lung injury) and calls for bans on nicotine flavors and vaping devices, Apple prohibited new vaping-related mobile apps from its iOS App Store in June 2019 and removed 181 previously approved vaping-related apps on November 15, 2019 [8,43]. Articles published on technology, vaping, and cannabis industry–affiliated websites in the following weeks decried this move by Apple, stating that the ban impacted device users' ability to safely use their products [44], removed a resource that assists people with quitting smoking [45], and would have detrimental effects on innovation [45], and that the app store approval process was inconsistent [46].

At a time of escalating concern about vaping, Apple's policy change may benefit public health by lessening youth exposure to content that glamorizes vaping. Exposure to vaping-related content online has been associated with greater intentions to vape [47] and greater likelihood of vaping [48] among youth. However, Apple's restrictions may also block important sources of information and vaping device control for adults seeking to use vaping devices safely.

The content and evidence base of vaping-related apps is currently unknown, though previous reviews and content analyses have examined apps related to smoking cessation, cannabis, alcohol [49], and other substance use [50]. Smoking cessation apps are abundant, though few have been demonstrated to be evidence-based [51-55], and many have limitations with usability [56]. A review [57] of cannabis apps conducted in 2014 found that the most common content areas largely consisted of cannabis strain classification guides, factoids about cannabis, and games, but few apps addressed negative health effects of cannabis use. To gain insight into the potential positive and negative effects of removing vaping-related apps from app stores, characterization of the remaining vaping-related apps in the Google Play Store, the main competing source for mobile apps, is an imperative first step.

Study Objective

We analyzed the top vaping-related mobile apps on the Google Play Store with respect to app characteristics, intended purpose, and provision of features and information for limiting the potential harms of smoking and vaping. To our knowledge, the Google Play Store did not enact any restrictions on vaping-related apps in response to EVALI. Within the context of government and industry regulatory policy changes surrounding vaping devices, understanding the types of apps available for mobile phones and tablets can shed light on how this form of digital media may reflect and influence information about nicotine and cannabis vaping available to consumers. Depending on their purposes and features, vaping-related apps

XSL•FO
RenderX

could encourage or discourage vaping. Some apps may provide resources for adults seeking to switch to vaping from smoking or to use vaping devices safely, while others may negatively impact youth by glamorizing vaping. As such, the removal of vaping-related apps may have mixed effects on population health, depending on which types of content predominate.

Methods

Selection of Apps

A search for apps using the keywords *vape* and *vaping* was performed on the Google Play Store on December 17, 2019 from an IP address in the United States. The names and order of the first 100 apps displayed in response to the search query were recorded. Similar to previous studies reviewing apps, only the first 100 apps of 250 search results were examined, because it is unlikely that individuals would browse further for a desired app without refining their search terms [56-58]. We chose not to include *e-cigarettes* or *e-cigs* as search terms because of their infrequent use outside of academic and regulatory discourses.

Resulting vaping apps available in the Google Play Store served as a proxy for apps that may have been available in the Apple App Store prior to November 2019, as a list of removed apps was not publicly available. Thus, coding vaping-related apps in the Google Play Store provided insight into the types of the apps both currently available to Google Play Store customers and previously available to Apple App Store customers. A search for *vaping* and *vape* in the Apple App Store in December 2019, performed on both a Mac and an iPhone, yielded zero results on the Mac Store and yielded only apps related to quitting vaping or to vapor-like image effects in the iPhone app store.

Rating of Apps

Three investigators (MM, EV, JT) each extracted app characteristics directly from the Google Play Store for one-third of the apps (33-34 apps per investigator). Extracted app characteristics included the developer name, content and age ratings, cost, average star ratings, number of reviewers, number of installations, last date updated, and URL. An app purpose coding guide assessed whether the app was relevant to vaping (yes or no), the intended substance (nicotine/tobacco, cannabis/THC, or unclear), and the purpose of the app.

The app purpose coding guide was initiated by MM and developed iteratively among the investigators. Any information on the app store webpage, including description and screenshots, was used for coding the app purpose and content. Apps could have multiple purposes. After initial coding, the investigators discussed tabulated results and aspects of the coding guide that were unclear, then refined the coding guide based on common patterns and any confusing aspects of the coding guide.

The second coding guide was then applied to 20% of the apps, which were triple coded. Each investigator coded 21 apps (7 recoded from their first pass, with 7 each from the 2 other coders). Unanimous agreement across all 3 investigators was then assessed for the app's relevance to vaping (yes or no), intended substance (nicotine/cannabis/unclear), and each of 10 potential purpose categories (coils, e-liquids, mods, shopping, games, social, device, quitting smoking, quitting vaping, other).

If an app was determined to be about something other than vaping, its purpose was not coded. Agreement before consensus was 19/21 (90%) for relevance to vaping, 12/17 (71%) intended substance, and 14-17/17 (82%-100%) for the purpose categories. Individual apps for which at least one investigator had a discordant code were discussed until consensus was reached.

A third and final coding guide was applied to the full list of apps. Of note, the coding guide was clarified so that apps referencing *heat-not-burn* were coded as vaping-related. While tobacco heat-not-burn devices are considered distinct from vaping devices, heat-not-burn of cannabis flower is often considered vaping, and neither involves combustion. Additionally, the group could not determine the purpose of 6 apps from the Google Play Store descriptions. These apps were downloaded and evaluated using additional information from the downloaded app itself.

The final 8 purpose categories were do-it-yourself (DIY) coils, DIY e-liquids, shopping, entertainment, social, device, quitting smoking, and quitting vaping. Summary percentages and means for app metadata and purposes were calculated.

Selection and Evaluation of Downloaded Apps

Next, apps within each of the 8 purpose categories were ranked by total number of installs, and the top 2 to 5 apps per category were downloaded for review. Instead of ranking the most popular apps, we ranked apps by popularity within categories, so that apps with less overall popularity but potentially important purposes would be included. The number of apps selected per category varied due to ties in the reported number of installs. Because apps could have multiple categories, this procedure resulted in a list of 18 apps with 10 to 1 million downloads each. Three of these apps disappeared from the store before they could be downloaded for review; one could be replaced with a premium version of the same app. A total of 16 apps were downloaded onto 2 Samsung Galaxy Tab A tablets and a Google Pixel 2 smartphone.

A random selection (6/16, 38%) was reviewed by all 3 investigators. Discrepancies were discussed, and the coding guide was updated. The final coding guide included evaluations of whether the content of the downloaded app matched the purpose category (yes or no) and whether it had the following types of information (yes or no): information about harms of vaping (eg, lung injury, nicotine dependence), information about safer vaping or DIY device use (eg, how to prevent explosions), and information about harms of smoking (eg, nicotine dependence, cardiovascular harms, cancer risk). For apps coded as being intended for quitting smoking or quitting vaping, the presence of a tracking feature in the downloaded app (eg, tracking days without smoking or vaping, tracking money saved) was noted (yes or no). For apps coded as pairing with devices, the presence of features for tracking temperature, dosage, or device locking was noted (yes or no). Differences between investigators were discussed until consensus was reached.

Finally, the Mobile App Rating Scale (MARS) was applied to all 16 downloaded apps. The MARS is 23-item multidimensional measure for rating the quality of mobile health apps, with 5 subscales in the areas of engagement (5 items), functionality (4

items), aesthetics (3 items), information (7 items, including affiliation of developer), and subjective rating (4 items) [59]. Subjective rating was not applied as these items involve hypothetical personal use. Each item was rated on a scale of 1 to 5, and ratings that differed by more than 2 points between investigators (eg, ratings of 1 vs 4 or 2 vs 5) were discussed. Investigators could adjust their ratings after discussion before averaging scores and did so 7 times across the 5 discussed apps. Some information items were rated as not applicable (N/A) and were excluded from average score calculations. For example, the item about meeting goals would be rated as N/A if the app purpose was not related to quitting smoking or vaping, and the visual information item would be rated as N/A if the app only contained text. Other smaller discrepancies were averaged without further discussion. Average ratings were calculated for each of the 4 subscales (engagement, functionality, aesthetics, information) and then averaged across the 3 investigator ratings. A final score was averaged for all downloaded apps and within each category.

Results

Overview of Vaping Apps

Of the top 100 apps captured by our search in December 2019, 79 (79%) were determined to be about vaping and were coded for purpose and other features. It was determined that 15 of the 21 apps not about vaping referred to cigarette smoking or only smoking cessation. An additional 6 apps could not be found again during content coding in January 2020 and February 2020 and were removed from the analysis. There were 4 sets of apps with both free and pro versions. The pro version typically cost money to download and did not have advertisements or had additional features. All app descriptions were in English, though some apps appeared to have content in other languages.

Of the 79 apps, most apps were free (60, 76%) or free with in-app purchases (13, 16%); 6 apps cost between US $0.99 and US $6.99 to download. Over half the apps had in-app advertisements (45, 57%). Only one-third of apps (36, 33%) were rated as Mature 17+, while the rest were rated as Teen (13, 16%) or Everyone (40, 51%) (Table 2).

There were 15 different Google Play Store–provided categories displayed with the app description, with the most common being tools and lifestyle, followed by health and fitness, simulation, and social. As one indicator of popularity, app downloads or installations ranged from at least 10 installs to over 1 million installs, with the first 10 apps presented in the search having at least 10,000 installs. As another set of popularity indicators, the average for the 68 apps with ratings was 4.0 stars, with a range of 2.5-5.0 stars by an average of 770 raters.

Investigator-coded app purposes are described in Table 2. The most common investigator-coded app purposes were creating DIY vaping e-liquids (28/79, 35%) and coils (25/79, 32%); 16 apps (16/79, 20%) were coded as for both creating coils and e-liquids. The next most common purposes were entertainment or games (19/79, 24%), social networking with other app users (16/79, 20%), and shopping (15/79, 19%). Social and shopping also tended to be co-occurring purposes (10/79, %). Finally, apps to help people quit smoking (8/79, 10%), directly control vaping devices (4/79, 5%), and quit vaping were less common (2/79, 3%). Out of 8 apps with quitting smoking features (typically a "cigarettes avoided" widget), 7 were also coded as DIY e-liquid or coil purposes. The majority referred to vaping with nicotine (43/79, 54%). Few apps were intended for cannabis (3/79, 4%), and 1 app referred to both nicotine and cannabidiol from cannabis (1/79, 1%). The intended substance was unclear in the remaining apps (31/79, 39%) (Table 3).

Because apps could fall into multiple categories, we did not statistically test differences in star ratings. However, we noted that the highest rated apps by users were in the DIY e-liquids and quitting smoking and vaping categories, while the lowest were in the device category. The most popular apps by number of installations were in the DIY e-liquids, DIY coils, and entertainment categories.

Table 2. Overview of vaping apps on Google Play Store in December 2019 (N=79).

Characteristic	Value, n (%)
Cost	
Price $0.99-$6.99	6 (8)
In-app purchases	13 (16)
Free	60 (76)
Advertisements	
Yes	45 (57)
No	34 (43)
Age rating	
Everyone	40 (51)
Teen	13 (16)
Mature 17+	26 (33)
Categories	
Tools	25 (32)
Lifestyle	22 (28)
Health and fitness	5 (6)
Simulation	4 (5)
Social	4 (5)
Art and design	4 (5)
Other[a]	15 (19)
Other content ratings	
Use of tobacco	7 (9)
Use of drugs	3 (4)
Drug reference	1 (1)
Tobacco reference	1 (1)
Language	1 (1)
Violence, blood	1 (1)
Total installs	
10 to 500	13 (16)
1000 to 5000	36 (46)
10,000 to 50,000	20 (25)
100,000	8 (10)
1,000,000	2 (3)
Ratings	
Stars out of 5 (n=68), mean (SD)	4 (0.7)
Raters (n=68), mean (SD)	770 (2607)
Last updated	
2013-2016	13 (16)
2017	11 (14)
2018	24 (30)
2019 (through November 14)	13 (16)
November 15, 2019-February 29, 2020	18 (23)

XSL•FO
RenderX

[a]Other includes adventure, education, entertainment, libraries and demo, maps and navigation, medical, music and audio, personalization, productivity, and travel and local categories.

Table 3. Purpose and intended substance ratings for vaping apps on Google Play Store (N=79).

Category	Description	Apps, n (%)	With star ratings, n	Star rating, mean (SD)	Installations, n
Purpose[a]					
DIY[b] e-liquids	Has features for creating e-liquids (eg, calculators with nicotine and propylene glycol or vegetable glycerin inputs, e-liquid recipes)	28 (35)	26	4.18 (0.60)	500+ to 1,000,000+
DIY coils	Has features for designing coils for DIY mods (eg, Ohm's law calculators)	25 (32)	25	4.01 (0.68)	500+ to 1,000,000+
Entertainment	Simulations, wallpapers, games	19 (24)	14	3.74 (0.65)	100+ to 1,000,000+
Social	Has features for connecting with other app users or people who vape, including the ability to review products	16 (20)	15	4.03 (0.79)	50+ to 100,000+
Shopping	For finding products or stores that sell vaping devices or products or e-liquids	15 (19)	13	3.92 (0.93)	50+ to 100,000+
Quitting smoking	For helping people quit smoking (using vaping or not); may include a tracking feature oriented around not using cigarettes	8 (10)	8	4.06 (0.52)	500+ to 10,000+
Device	Pairs with and has features for modifying a device	4 (5)	2	2.65 (0.07)	100+ to 10,000+
Quitting vaping	For helping people quit vaping	2 (3)	1	4.10 (N/A[c])	10+ to 1000+
Substance					
Cannabis	N/A	3 (4)	2	3.55 (1.34)	100+ to 10,000+
Nicotine	N/A	43 (54)	38	3.97 (0.62)	10+ to 1,000,000+
Unclear	N/A	31 (39)	26	4.07 (0.78)	50+ to 100,000+
Both	N/A	1 (1)	1	4.70 (N/A)	10,000+

[a]Apps could have multiple purposes.

[b]DIY: do-it-yourself.

[c]N/A: not applicable.

Downloaded Apps

The majority of downloaded apps (14/16, 88%) matched the descriptions in the app store (Table 4). The exceptions were that 1 app coded as having quitting smoking features did not have any such features, and 1 app that was coded as not having shopping features did, in fact, have links to shopping through the app. Few apps had information about harms of vaping (4/16, 25%), safer vaping (3/16, 19%), or harms of smoking (2/16, 13%). When they did have such information, it was often difficult to find. Of note, 2 of the entertainment apps had a feature where an avatar would cough when vaping too much, which may normalize moderation in vaping; 5 apps had tracking features, which mainly recorded days passed since a user-provided quit-smoking date; and 2 of the apps with tracking features also displayed money saved and health benefits. Both device apps appeared to have temperature controls, but did not have dosage or locking settings visible, though these may have become apparent once paired with a device.

Overall, the 16 downloaded apps had a mean MARS score of 3.63, with a highest mean subscore for functionality (MARS score: mean 4.13) and lowest mean subscore for engagement (MARS score: mean 3.36). Within the subtypes of purposes, the highest mean MARS scores were for social, shopping, and device apps, and the lowest mean scores were for the quitting smoking and DIY coils apps.

XSL•FO
RenderX

Table 4. Ratings of downloaded vaping apps (N=16).

Characteristic and subcharacteristic	Value
Purpose, n (%)	
Matches coded purpose	14 (88)
Information, n (%)	
Harms of vaping	4 (25)
Safer vaping	3 (19)
Harms of smoking	2 (13)
Behavior change, n (%)	
Tracking feature	5 (31)
Device (n=2), n (%)	
Temperature	2 (100)
Dosage	0 (0)
Locking	0 (0)
MARS[a] (out of 5, all downloaded apps), mean (range)	
Engagement	3.36 (2.80-4.20)
Functionality	4.13 (3.08-4.58)
Aesthetics	3.40 (2.22-4.50)
Information	3.61 (2.50-4.83)
Summary	3.63 (2.77-4.47)
MARS (out of 5, by purpose category)[b], mean (range)	
DIY[c] e-liquids (n=5)	3.54 (3.16-4.04)
DIY coils (n=3)	3.44 (3.16-3.89)
Entertainment (n=3)	3.66 (3.00-4.47)
Social (n=3)	4.22 (4.04-4.47)
Shopping (n=2)	4.09 (4.04-4.15)
Quitting smoking (n=2)	3.08 (2.89-3.49)
Device (n=2)	4.04 (3.93-4.15)
Quitting vaping (n=2)	3.93 (3.80-4.06)

[a]MARS: Mobile App Rating Scale.

[b]Top 2 to 5 apps per category by number of installations; categories with 3 to 5 apps had ties in the number of installations.

[c]DIY: do-it-yourself.

Discussion

This study examined the content of the first 100 mobile apps on the Google Play Store using *vaping* and *vape* as search terms 1 month after Apple's ban on vaping-related apps, which was enacted in response to concerns about youth nicotine vaping and EVALI (a lung injury syndrome linked to an additive to cannabis vaping products). Of 79 apps determined to be related to vaping, over half were related to nicotine, while only a few were for cannabis, and the rest were unclear in intended substance. The most popular app content, with respect to both number of installations (several with over 1 million) and percentage of these 79 vaping apps, was creating DIY liquids (28/79, 35%) and DIY coils (25/79, 32%), with 20% (16/79) in both categories (37 total). This may reflect a strong interest in

DIY hobbies in vaping culture. DIY may allow users to control the customization process, play with novelty, save money, and achieve higher nicotine concentrations [60,61]. Overall, the main purposes of the majority of these vaping-related apps on the Google Play Store were to help people continue to vape nicotine.

Apps that had features to support quitting smoking or vaping were relatively rare (8/79,10%; 2/79, 3%; respectively). Most apps that supported smoking cessation encouraged users to quit smoking by switching to vaping. These apps also contained other features to promote or facilitate vaping, such as e-liquid recipes. The 2 apps for quitting vaping that were downloaded had above average MARS scores but relatively few installations, while the 3 apps for quitting smoking that were downloaded had below average MARS scores and low subscores on

aesthetics and information. This points to the need for apps to promote vaping cessation using evidence-based behavior-change information and strategies, along with engaging usable interfaces.

There were also few apps that paired with devices (4/79, 5%), and these device apps were mainly for cannabis. Devices that pair with apps using Bluetooth technology are more expensive, which may explain their lower popularity in terms of installs. Features available in the downloaded device apps indicated that the user could control temperature [62], which may limit throat and lung damage, but not dosage, and there were no locking controls for users with children or those wishing to moderate their use.

Given current concerns about youth vaping, the findings that over half the apps had no age controls and that a large proportion of apps without age controls was in the DIY categories (24/37, 65%) are especially concerning. A smaller proportion of entertainment apps had no age controls (5/19, 26%), though many of the apps with age controls were set at Teen. Age controls may allow parents who utilize family controls to restrict their children's access to these apps.

Of the popular apps that were downloaded and reviewed in-depth, few apps presented information about the harms of vaping or smoking or included information about safer vaping. Information about harms of smoking consisted of articles comparing the harms of smoking (eg, combustion and cancer-causing ingredients) to vaping. One app had a widget for tracking "days without smoking" that included "gained days of life" and "avoided radiographs" calculations.

Information presented about harms of vaping included acknowledgments of the importance of moderating nicotine intake, the addictiveness of nicotine, or harms of vaping in front of children. The 2 downloaded apps that were intended to help people quit vaping both included links to news stories about young adults with EVALI and articles about concerns with youth vaping and the intentions of vaping companies. Safer vaping information included recommendations about coil and battery materials, causes of e-cigarette explosions, and cautions against mixing e-liquids incorrectly. It should be noted that in 2 entertainment apps with simulated vaping games, the vaping avatar would cough audibly when they "inhaled" a large amount of vapor, which could be seen as encouraging moderation in use. Several of these apps included a tracking feature that displayed the number of days since quitting smoking and the number of cigarettes avoided. Self-monitoring is an important component of a smoking cessation plan but is likely insufficient by itself [63]. In addition to a need for apps on the Google Play

store that assist people with quitting vaping, study results indicate a need for informational apps to better describe the pros and cons of vaping.

Assuming that these Google Play Store apps were similar in content to the Apple App Store apps that were removed, it appears that Apple's ban would have had a minimal effect on people who vape with the intention of quitting smoking or who are seeking information about safer vaping. Nevertheless, the decision to remove the vaping-related apps appears to have been taken by Apple in response to rising EVALI cases, which were primarily attributed to cannabis oil additives, rather than nicotine liquids [33,34]. There appeared to be little publicly available information detailing how apps were determined to be removed, echoing other calls for increased transparency and additional research regarding allowed app platform content and other issues like privacy [39,64]. Future research should explore other cases touching on the who and how of regulation of apps related to controversial health-behavior for which there is not yet a consensus among health experts. Future research should also examine more explicitly the relationship between vaping app use and vaping behaviors.

There were several limitations to this study. First, not all apps that came up in the initial search were reviewed, though most app users would likely not browse more than 100 apps without refining their search. Additionally, the app store gets updated continuously, and a search on a different date may present different results. Indeed, several apps were no longer available a few weeks after the initial search. While the number of installations was recorded for each as a signal of popularity, people may download an app and not use it at all or only use it a limited number of times. Only 1 of the reviewed apps was also available on the iPhone App Store, but it is unclear which of the apps we reviewed were removed from or denied approval in the Apple App Store. Although we only coded Google Play Store apps, a search for *vaping* in the mobile Apple App Store in December 2019 confirmed that the remaining apps were related to quitting smoking or quitting vaping or were unrelated to vaping behavior.

Based on this review of vaping-related apps in the Google Play Store, it appears that the Apple vaping app ban would have had a minimal effect on adults seeking to switch away from smoking or seeking to vape more safely. Most vaping-related apps in the Google Play Store were for purposes related to continuing vaping and had limited age-based access restrictions. Few apps were for controlling device settings, assisting with quitting smoking or vaping, or disseminating information about safer vaping.

Conflicts of Interest

None declared.

References

1. Walley SC, Wilson KM, Winickoff JP, Groner J. A Public Health Crisis: Electronic Cigarettes, Vape, and JUUL. Pediatrics 2019 Jun 23;143(6):e20182741. [doi: 10.1542/peds.2018-2741] [Medline: 31122947]

2.   Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction 2015 Dec 12;110(11):1699-1704 [FREE Full text] [doi: 10.1111/add.13036] [Medline: 26264448]

3.   King BA, Jones CM, Baldwin GT, Briss PA. The EVALI and Youth Vaping Epidemics — Implications for Public Health. N Engl J Med 2020 Feb 20;382(8):689-691. [doi: 10.1056/nejmp1916171]

4.   Erku D, Gartner CE, Morphett K, Snoswell CL, Steadman KJ. Nicotine vaping products as a harm reduction tool among smokers: Review of evidence and implications for pharmacy practice. Res Social Adm Pharm 2020 Oct;16(9):1272-1278. [doi: 10.1016/j.sapharm.2020.02.002] [Medline: 32061550]

5.   Gordon T, Fine J. Cornering the Suspects in Vaping-Associated EVALI. N Engl J Med 2020 Feb 20;382(8):755-756. [doi: 10.1056/nejme2001065]

6.   Miech R, Johnston L, O'Malley PM, Bachman JG, Patrick ME. Adolescent Vaping and Nicotine Use in 2017–2018 — U.S. National Estimates. N Engl J Med 2019 Jan 10;380(2):192-193. [doi: 10.1056/nejmc1814130]

7.   Youth Tobacco Use: Results from the National Youth Tobacco Survey. Food And Drug Administration. 2019. URL: https://www.fda.gov/tobacco-products/youth-and-tobacco/youth-tobacco-use-results-national-youth-tobacco-survey [accessed 2020-05-06]

8.   Fried I, Allen M. Exclusive: Apple to remove vaping apps from store. Axios. 2019. URL: https://www.axios.com/exclusive-apple-to-remove-vaping-apps-from-store-8669fd94-e92a-4ce4-a9e2-ce5afa598b67.html [accessed 2020-05-06]

9.   Nutt DJ, Phillips LD, Balfour D, Curran HV, Dockrell M, Foulds J, et al. Estimating the harms of nicotine-containing products using the MCDA approach. Eur Addict Res 2014 Apr 3;20(5):218-225 [FREE Full text] [doi: 10.1159/000360220] [Medline: 24714502]

10.   Ghosh S, Drummond M. Electronic cigarettes as smoking cessation tool: are we there? Curr Opin Pulm Med 2017;23(2):0348. [doi: 10.1097/mcp.0000000000000348]

11.   Hajek P, Phillips-Waller A, Przulj D, Pesola F, Myers Smith K, Bisal N, et al. A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy. N Engl J Med 2019 Feb 14;380(7):629-637. [doi: 10.1056/nejmoa1808779]

12.   Singh S, Windle SB, Filion KB, Thombs BD, O'Loughlin JL, Grad R, et al. E-cigarettes and youth: Patterns of use, potential harms, and recommendations. Prev Med 2020 Mar 03;133:106009. [doi: 10.1016/j.ypmed.2020.106009] [Medline: 32027913]

13.   National Academy of Science Engineering and Medicine. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.

14.   Azagba S, Manzione L, Shan L, King J. Trends in Smoking Behaviors Among US Adolescent Cigarette Smokers. Pediatrics 2020 Mar 03;145(3):e20193047. [doi: 10.1542/peds.2019-3047] [Medline: 32015140]

15.   Miech R, Schulenberg J, Johnston L. National adolescent drug trends in 2018. In: Monitoring the Future. Ann Arbor, MI: University of Michigan; 2019.

16.   Boykan R, Goniewicz ML, Messina CR. Evidence of Nicotine Dependence in Adolescents Who Use Juul and Similar Pod Devices. Int J Environ Res Public Health 2019 Jun 17;16(12):2135 [FREE Full text] [doi: 10.3390/ijerph16122135] [Medline: 31212888]

17.   2016 Surgeon General's Report: Cigarette Use Among Youth and Young Adults. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/tobacco/data_statistics/sgr/e-cigarettes/index.htm [accessed 2020-05-06]

18.   Gravely S, Driezen P, Ouimet J, Quah ACK, Cummings KM, Thompson ME, et al. Prevalence of awareness, ever-use and current use of nicotine vaping products (NVPs) among adult current smokers and ex-smokers in 14 countries with differing regulations on sales and marketing of NVPs: cross-sectional findings from the ITC Project. Addiction 2019 Jun 20;114(6):1060-1073 [FREE Full text] [doi: 10.1111/add.14558] [Medline: 30681215]

19.   Erku DA, Kisely S, Morphett K, Steadman KJ, Gartner CE. Framing and scientific uncertainty in nicotine vaping product regulation: An examination of competing narratives among health and medical organisations in the UK, Australia and New Zealand. Int J Drug Policy 2020 Apr;78:102699. [doi: 10.1016/j.drugpo.2020.102699] [Medline: 32086156]

20.   Center for Tobacco-Free Kids. 2020. URL: https://www.tobaccofreekids.org/assets/factsheets/0398.pdf [accessed 2020-05-06]

21.   Fischer B, Russell C, Sabioni P, van den Brink W, Le Foll B, Hall W, et al. Lower-Risk Cannabis Use Guidelines: A Comprehensive Update of Evidence and Recommendations. Am J Public Health 2017 Aug;107(8):e1-e12. [doi: 10.2105/ajph.2017.303818]

22.   Aston ER, Scott B, Farris SG. A qualitative analysis of cannabis vaporization among medical users. Exp Clin Psychopharmacol 2019 Aug;27(4):301-308 [FREE Full text] [doi: 10.1037/pha0000279] [Medline: 31120278]

23.   Tashkin DP. How beneficial is vaping cannabis to respiratory health compared to smoking? Addiction 2015 Dec 16;110(11):1706-1707. [doi: 10.1111/add.13075] [Medline: 26471152]

24.   Borodovsky JT, Crosier BS, Lee DC, Sargent JD, Budney AJ. Smoking, vaping, eating: Is legalization impacting the way people use cannabis? Int J Drug Policy 2016 Oct;36:141-147 [FREE Full text] [doi: 10.1016/j.drugpo.2016.02.022] [Medline: 26992484]

25.   Schauer GL, Njai R, Grant-Lenzy AM. Modes of marijuana use - smoking, vaping, eating, and dabbing: Results from the 2016 BRFSS in 12 States. Drug Alcohol Depend 2020 Apr 01;209:107900. [doi: 10.1016/j.drugalcdep.2020.107900] [Medline: 32061947]

XSL•FO
RenderX

26.   Meacham MC, Paul MJ, Ramo DE. Understanding emerging forms of cannabis use through an online cannabis community: An analysis of relative post volume and subjective highness ratings. Drug Alcohol Depend 2018 Jul 01;188:364-369 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.03.041] [Medline: 29883950]

27.   Giroud C, de Cesare M, Berthet A, Varlet V, Concha-Lozano N, Favrat B. E-Cigarettes: A Review of New Trends in Cannabis Use. Int J Environ Res Public Health 2015 Aug 21;12(8):9988-10008 [FREE Full text] [doi: 10.3390/ijerph120809988] [Medline: 26308021]

28.   Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in Internet Searches for Cannabidiol (CBD) in the United States. JAMA Netw Open 2019 Oct 02;2(10):e1913853 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.13853] [Medline: 31642924]

29.   Vandrey R, Raber JC, Raber ME, Douglass B, Miller C, Bonn-Miller MO. Cannabinoid Dose and Label Accuracy in Edible Medical Cannabis Products. JAMA 2015 Jun 23;313(24):2491-2493. [doi: 10.1001/jama.2015.6613] [Medline: 26103034]

30.   Bonn-Miller MO, Loflin MJE, Thomas BF, Marcu JP, Hyke T, Vandrey R. Labeling Accuracy of Cannabidiol Extracts Sold Online. JAMA 2017 Dec 07;318(17):1708-1709 [FREE Full text] [doi: 10.1001/jama.2017.11909] [Medline: 29114823]

31.   Friedman AS. Association of vaping-related lung injuries with rates of e-cigarette and cannabis use across US states. Addiction 2020 Aug 25:00-00. [doi: 10.1111/add.15235] [Medline: 32840932]

32.   Ghinai I, Navon L, Gunn JK, Duca LM, Brister S, Love S, et al. Characteristics of Persons Who Report Using Only Nicotine-Containing Products Among Interviewed Patients with E-cigarette, or Vaping, Product Use-Associated Lung Injury - Illinois, August-December 2019. MMWR Morb Mortal Wkly Rep 2020 Jan 24;69(3):84-89 [FREE Full text] [doi: 10.15585/mmwr.mm6903e1] [Medline: 31971930]

33.   Krishnasamy VP, Hallowell BD, Ko JY, Board A, Hartnett KP, Salvatore PP, Lung Injury Response Epidemiology/Surveillance Task Force. Update: Characteristics of a Nationwide Outbreak of E-cigarette, or Vaping, Product Use-Associated Lung Injury - United States, August 2019-January 2020. MMWR Morb Mortal Wkly Rep 2020 Jan 24;69(3):90-94 [FREE Full text] [doi: 10.15585/mmwr.mm6903e2] [Medline: 31971931]

34.   Hall W, Gartner C, Bonevski B. Lessons from the public health responses to the US outbreak of vaping-related lung injury. Addiction 2020 May 04:00-00. [doi: 10.1111/add.15108] [Medline: 32364274]

35.   Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]

36.   Fox S, Duggan M. Mobile Health 2012. Pew Research Center. 2012. URL: https://www.pewresearch.org/internet/2012/11/08/mobile-health-2012/ [accessed 2020-05-06]

37.   Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]

38.   Clement J. Number of apps available in leading app stores as of 1st quarter 2020. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/#statisticContainer [accessed 2020-08-25]

39.   Hestres LE. App Neutrality: Apple's App Store and Freedom of Expression Online. SSRN Journal 2013;7:1265-1280. [doi: 10.2139/ssrn.2032050]

40.   Viswanathan P. iOS App Store vs Google Play Store: Which is better for app developers? Lifewire. 2020 Mar 09. URL: https://www.lifewire.com/ios-app-store-vs-google-play-store-for-app-developers-2373130 [accessed 2020-05-07]

41.   Restricted Content Developer Policy Center. Google Play. 2020. URL: https://play.google.com/about/restricted-content/ [accessed 2020-05-11]

42.   Apple. App Store Review Guidelines. URL: https://developer.apple.com/app-store/review/guidelines/#physical-harm [accessed 2020-05-10]

43.   Apple removes vaping apps from app store. BBC. 2019 Nov 15. URL: https://www.bbc.com/news/technology-50434755 [accessed 2020-05-07]

44.   Stickstones J. What Apple's Vape App Ban Means To You. Planet of the Vapes. 2019 Nov 22. URL: https://www.planetofthevapes.com/blogs/blog/what-apples-vape-app-ban-means-to-you [accessed 2020-05-07]

45.   Medal A. Assessing the Effects of Apple's Ban on Vaping Apps. Entrepreneur. 2019 Nov 27. URL: https://www.entrepreneur.com/article/342899 [accessed 2020-05-07]

46.   McDonald J. Apple Removes Vaping Content from App Store. Vaping360. 2019 Nov 16. URL: https://vaping360.com/vape-news/86676/apple-removes-vaping-apps-from-the-app-store/ [accessed 2020-05-07]

47.   Vogel E, Ramo D, Rubinstein M, Delucchi KL, Darrow S, Costello C, et al. Effects of Social Media on Adolescents' Willingness and Intention to Use E-Cigarettes: An Experimental Investigation. Nicotine Tob Res 2020 Jan 08:00-00. [doi: 10.1093/ntr/ntaa003] [Medline: 31912147]

48.   Camenga D, Gutierrez K, Kong G, Cavallo D, Simon P, Krishnan-Sarin S. E-cigarette advertising exposure in e-cigarette naïve adolescents and subsequent e-cigarette use: A longitudinal cohort study. Addict Behav 2018 Jun;81:78-83 [FREE Full text] [doi: 10.1016/j.addbeh.2018.02.008] [Medline: 29432916]

XSL•FO

RenderX

49.   Weaver ER, Horyniak DR, Jenkinson R, Dietze P, Lim MS. "Let's get Wasted!" and Other Apps: Characteristics, Acceptability, and Use of Alcohol-Related Smartphone Applications. JMIR Mhealth Uhealth 2013 Jul 25;1(1):e9 [FREE Full text] [doi: 10.2196/mhealth.2709] [Medline: 25100681]

50.   Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone Apps Targeting Alcohol and Illicit Substance Use: Systematic Search in in Commercial App Stores and Critical Content Analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]

51.   Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun 19;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]

52.   Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec 27;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]

53.   Regmi D, Tobutt C, Shaban S. Quality and Use of Free Smoking Cessation Apps for Smartphones. Int J Technol Assess Health Care 2018 Sep 18;34(5):476-480. [doi: 10.1017/s0266462318000521]

54.   Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile Applications for the Treatment of Tobacco Use and Dependence. Curr Addict Rep 2019 Jul 9;6(2):86-97 [FREE Full text] [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]

55.   Rajani NB, Weth D, Mastellos N, Filippidis FT. Use of gamification strategies and tactics in mobile applications for smoking cessation: a review of the UK mobile app market. BMJ Open 2019 Jun 17;9(6):e027883. [doi: 10.1136/bmjopen-2018-027883] [Medline: 31213452]

56.   Ferron JC, Brunette MF, Geiger P, Marsch LA, Adachi-Mejia AM, Bartels SJ. Mobile Phone Apps for Smoking Cessation: Quality and Usability Among Smokers With Psychosis. JMIR Hum Factors 2017 Mar 03;4(1):e7 [FREE Full text] [doi: 10.2196/humanfactors.5933] [Medline: 28258047]

57.   Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis Mobile Apps: A Content Analysis. JMIR Mhealth Uhealth 2015 Aug 12;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]

58.   Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]

59.   Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

60.   Cox S, Leigh NJ, Vanderbush TS, Choo E, Goniewicz ML, Dawkins L. An exploration into "do-it-yourself" (DIY) e-liquid mixing: Users' motivations, practices and product laboratory analysis. Addict Behav Rep 2019 Jul;9:100151 [FREE Full text] [doi: 10.1016/j.abrep.2018.100151] [Medline: 31193745]

61.   Davis B, Dang M, Kim J, Talbot P. Nicotine concentrations in electronic cigarette refill and do-it-yourself fluids. Nicotine Tob Res 2015 Mar 26;17(2):134-141 [FREE Full text] [doi: 10.1093/ntr/ntu080] [Medline: 24862971]

62.   Lanz C, Mattsson J, Soydaner U, Brenneisen R. Medicinal Cannabis: In Vitro Validation of Vaporizers for the Smoke-Free Inhalation of Cannabis. PLoS One 2016 Jan 19;11(1):e0147286 [FREE Full text] [doi: 10.1371/journal.pone.0147286] [Medline: 26784441]

63.   Clinical Practice Guideline Treating Tobacco Use Dependence 2008 Update Panel, Liaisons, Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]

64.   Shilton K, Greene D. Linking Platforms, Practices, and Developer Ethics: Levers for Privacy Discourse in Mobile Application Development. J Bus Ethics 2017 Mar 28;155(1):131-146. [doi: 10.1007/s10551-017-3504-8]

Abbreviations

DIY: do-it-yourself
EVALI: electronic or vaping product use associated lung injury
IP: internet protocol
MARS: Mobile App Rating Scale
THC: tetrahydrocannabinol
URL: uniform resource locator

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

Self-Monitoring App Preferences for Sun Protection: Discrete Choice Experiment Survey Analysis

Vasileios Nittas[1], BSc, MSc; Margot Mütsch[1], MPH, MSc, PhD; Julia Braun[1], MSc, PhD; Milo Alan Puhan[1], MD, PhD

Epidemiology, Biostatistics, and Prevention Institute, University of Zurich, Zurich, Switzerland

Corresponding Author:
Vasileios Nittas, BSc, MSc
Epidemiology, Biostatistics, and Prevention Institute
University of Zurich
Hirschengraben 84
Zurich, 8001
Switzerland
Phone: 41 44 63 44946
Email: vasileios.nittas@uzh.ch

Abstract

Background: The availability and use of health apps continues to increase, revolutionizing the way mobile health interventions are delivered. Apps are increasingly used to prevent disease, improve well-being, and promote healthy behavior. On a similar rise is the incidence of skin cancers. Much of the underlying risk can be prevented through behavior change and adequate sun protection. Self-monitoring apps have the potential to facilitate prevention by measuring risk (eg, sun intensity) and encouraging protective behavior (eg, seeking shade).

Objective: Our aim was to assess health care consumer preferences for sun protection with a self-monitoring app that tracks the duration and intensity of sun exposure and provides feedback on when and how to protect the skin.

Methods: We conducted an unlabeled discrete choice experiment with 8 unique choice tasks, in which participants chose among 2 app alternatives, consisting of 5 preidentified 2-level attributes (self-monitoring method, privacy control, data sharing with health care provides, reminder customizability, and costs) that were the result of a multistep and multistakeholder qualitative approach. Participant preferences, and thus, the relative importance of attributes and their levels were estimated using conditional logit modeling. Analyses consisted of 200 usable surveys, yielding 3196 observations.

Results: Our respondents strongly preferred automatic over manually operated self-monitoring (odds ratio [OR] 2.37, 95% CI 2.06-2.72) and no cost over a single payment of 3 Swiss francs (OR 1.72, 95% CI 1.49-1.99). They also preferred having over not having the option of sharing their data with a health care provider of their choice (OR 1.66, 95% CI 1.40-1.97), repeated over single user consents, whenever app data are shared with commercial thirds (OR 1.57, 95% CI 1.31-1.88), and customizable over noncustomizable reminders (OR 1.30, 95% CI 1.09-1.54). While most participants favored thorough privacy infrastructures, the attribute of privacy control was a relatively weak driver of app choice. The attribute of self-monitoring method significantly interacted with gender and perceived personal usefulness of health apps, suggesting that female gender and lower perceived usefulness are associated with relatively weaker preferences for automatic self-monitoring.

Conclusions: Based on the preferences of our respondents, we found that the utility of a self-monitoring sun protection app can be increased if the app is simple and adjustable; requires minimal effort, time, or expense; and has an interoperable design and thorough privacy infrastructure. Similar features might be desirable for preventive health apps in other areas, paving the way for future discrete choice experiments. Nonetheless, to fully understand these preference dynamics, further qualitative or mixed method research on mobile self-monitoring-based sun protection and broader preventive mobile self-monitoring is required.

International Registered Report Identifier (IRRID): RR2-10.2196/16087

Introduction

Background

The global increase in smartphone ownership is unprecedented [1]. Worldwide, approximately seven billion people are estimated to own a mobile phone, more than half of which are smartphones [1,2]. Simultaneously, increased access to the internet and rising advocacy for person-centered care promote a rapidly growing mobile health (mHealth) market [3]. Inevitably, the availability of free or low-cost health apps has increased, surpassing 300,000 as of 2017 [3,4]. Shifting beyond mere SMS, the multifunctionality of apps has the potential to revolutionize the way mHealth interventions are delivered [5]. They can be used to prevent disease and facilitate well-being, such as monitoring behavior and promoting healthy life-styles; to manage existing conditions, such as providing behavioral therapies and assisting medication adherence; and to support rehabilitation and health care access [5-7].

Mobile Self-Monitoring Apps for Sun Protection

Skin diseases, such as skin cancers, that are caused by sun exposure are also on the rise [8]. Primarily prevalent in Caucasian, fair-skinned populations, melanoma and nonmelanoma skin cancers increasingly affect younger age groups [8]. Beyond uncontrollable risk factors such as genetics, the most significant and avoidable risk factor is exposure to ultraviolet light (eg, sun, tanning beds) [8]. Targeted behavior changes, such as improved sun protection (eg, shade, protective clothing, sunscreen use during exposure to the sun) as well as full avoidance of tanning beds, are key elements to the mitigation of a growing public health burden [8].

Primary prevention (eg, wellness, physical activity, diet, healthy behaviors) accounts for a large proportion of health apps [5]. Most of these have functions to collect and analyze health-related data, enabling simple and continuous self-monitoring [5]. These trends lead to an ever-growing volume of electronic patient-generated health data, defined as nonclinical health information [9]. Using mHealth apps to self-monitor is expected to facilitate prevention by measuring risk exposure (eg, sun intensity), which in turn can increase risk awareness and encourage healthy behavior (eg, seeking shade, applying sunscreen) [5,10]. Acknowledging the reach and potential of self-monitoring apps and the increasing burden of preventable skin cancers, efforts to understand how the former are to be designed and utilized to successfully prevent the latter are timelier than ever before.

Consumer Preferences for Sun Protection With Self-Monitoring Apps

Nonetheless, health apps are often deleted or abandoned rapidly as they fail to meet user needs and expectations [11]. One approach to meeting these is by exploring the preferences of prospective users. This is particularly important for health apps that tend to have a niche and often periodic use, such as in the case of sun protection apps. Within often limited time windows, such apps need to deliver a certain value in order to be considered for download and adequate use. In addition, both self-monitoring as well as disease prevention predominantly depend on individual motivation (acceptance and engagement), which is in turn linked to individual preferences [12-14]. Despite that, consumer preferences for self-monitoring-based sun protection apps have not yet been adequately explored. We aim to fill this gap by reporting the findings on the preferences of 200 health care consumers.

Aims

In the context of sun protection, this study aimed to assess which attributes of a theoretical self-monitoring app are perceived as more or less important by health care consumers. This was guided by the following objectives: (1) to elicit how consumer preferences are distributed among a set of preidentified app attributes and (2) to assess whether preferences vary across individual characteristics, such as age, gender, education, perceived health and perceived health app usefulness.

Methods

Study Design

Consumer preferences were elicited with a self-administered discrete choice experiment survey. Discrete choice experiments provide respondents with 2 or more alternative options of a product or service (in our case, 2 alternatives of the same app), asking them to repeatedly choose the most desired one [15,16]. All alternatives are defined by the same attributes, however, contain different manifestations (levels) of these. Every attribute included in a discrete choice experiment takes 2 or more forms. For example, if cost is included as an attribute, its levels could take 2 (or more) forms, such as free or 3 Swiss francs. These repeated choices allow for statistically eliciting which of those attributes (and their levels) are perceived as more valuable, and thus, yield the highest utility [16]. A person's perceived utility of an attribute depends on the attribute's level (eg, for the cost attribute, a less expensive app might be perceived as more valuable than a more expensive app). While traditionally rooted in economic research, discrete choice experiments are gradually gaining popularity in digital health research [15]. Our methodology follows the guidelines reported by the ISPOR Good Research Practices for Conjoint Analysis Task Force [17]. Our study design followed these 5 steps: (1) qualitative work for the identification of key attributes, (2) survey and scenario development, (3) piloting and survey adjustments, (4) survey administration and data collection, and (5) choice data analysis. A detailed description of the applied methodology, as well as a justification of our methodological rationale have been reported in a previously published registered protocol [18].

Attribute Identification

We followed good practice standards and identified the attributes of our discrete choice experiment by (1) using the results of the previously conducted systematic scoping review, (2) conducting a rapid review of systematic reviews on the use of self-monitoring data for primary prevention [19], (3) completing 13 semistructured expert interviews and 12 health care consumer interviews as well as (4) 2 internal review rounds (Figure 1) [18].

Figure 1. Attribute selection process. DCE: discrete choice experiment.
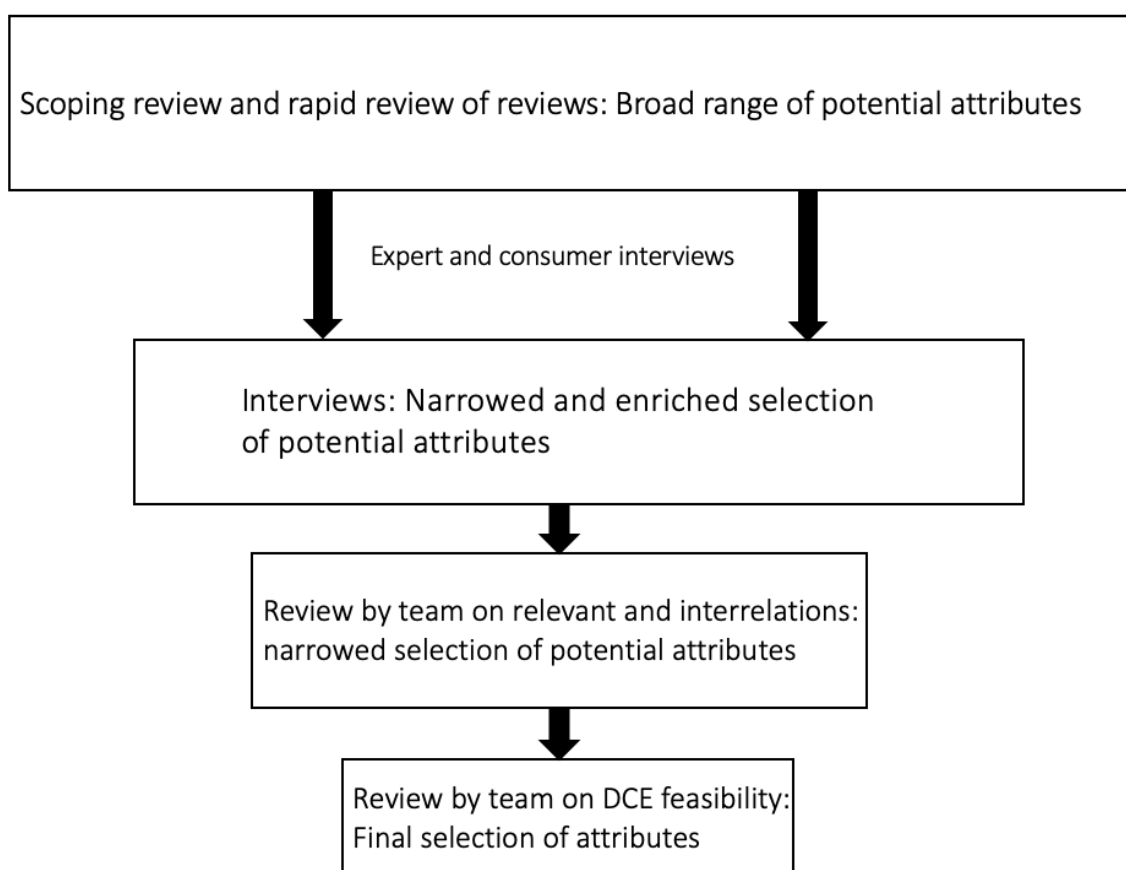
We selected experts that covered all relevant expertise fields, including digital self-monitoring, digital prevention, eHealth ethics, clinical science, and citizen science [18]. Health care consumer interviews were gender balanced and consisted of adults. Interviews were recorded, transcribed, and analyzed following a mixed deductive and inductive approach. The literature reviews led to a broad range of potential attributes, which was narrowed down and complemented by the interviews. This was followed by 2 internal review rounds, assessing the relevance, interrelations, and feasibility of identified attributes, leading to the selection and inclusion of the following 5 attributes: (1) method of self-monitoring, (2) privacy control, (3) data sharing with a health care provider, (4) reminder customizability, and (5) costs. An overview of included attributes with their levels and definitions is provided in Table 1.

For each attribute, we chose 2 realistic, relatable, and easy-to-understand levels, primarily based on our literature and interview findings. Attributes and levels were purposively kept simple and unambiguous, ensuring that all respondents could complete the survey with minimum effort. Long or too complex discrete choice experiments can pose a disproportionate cognitive burden on respondents, which increases the risk of incomplete or inaccurate responses. We hypothesize that health care consumers will prefer automatic over manual self-monitoring, customizable over noncustomizable reminders, and no cost over a single payment have all been confirmed. For the other two attributes, a hypothesis cannot be clearly formulated.

Table 1. Identified attributes and attribute levels.

Attributes	Descriptions and attribute levels
Self-monitoring method	**How would you like your data to be collected?**
	No manual entry (automatic entry)
	Manual entry once a day
Privacy control	**If your data are being shared with third commercial parties, how would you like to control when and with whom your data are shared?**
	I will only receive information on potential receivers once and will be asked to provide informed consent once
	I will be informed and provide consent whenever my data are provided to third parties
Data sharing with health care provider	**Would you like to share the data collected with your health care provider, to be discussed at your next visit?**
	Yes
	No
Reminder customizability	**How would you prefer the times and frequency of your reminders to be set?**
	I set the time and frequency of my reminders myself (customizable)
	The app sets the times and frequency of reminders automatically, based on my data
Costs	**Are there any costs associated with downloading the app, and if yes, how high are these?**
	Free
	One-time 3 Swiss francs

Survey Design and Scenario

Our 5 attributes, each with 2 levels, would lead to a full factorial design of 32 ($2^5 = 32$) possible choice sets [20]. However, asking participants to make 32 discrete choices would be highly impractical, due to limited time and high complexity. Hence, we followed common practice and reduced the survey to a subset of possible choice sets, known as fractional factorial design.

Based on D-optimality criteria, the resulting fractional factorial design consisted of 8 choice sets, each providing 2 app alternatives [21,22]. In addition to the 8 fractional factorial design choice sets, we decided to include a ninth set, which was used to assess the quality of responses. The ninth choice set was identical to a previous one, allowing us to assess whether responses were consistent (same app chosen for both identical sets) or not (different app chosen between the 2 identical sets). An example choice set is provided in Figure 2.

Figure 2. Choice set.

Please tick your preferred option	App A ☐	App B ☐
Self-monitoring method	**automatic** self-monitoring (the app collects data alone)	**manual** self-monitoring (I have to enter data into the app)
Sharing of data to commercial third (private entities)	I am asked for consent **once**, before downloading the app	I am asked for consent **every time** my data are shared
Sharing of data with a healthcare provider (eg, your GP)	**No**, the option is not there	**Yes**, the option is there
Customizability of reminders	I can **customize** the reminders I receive from the app on my smartphone	The app sends reminders in an automatic manner (**not customizable**)
Price	Free	One-time payment of 3 Swiss Francs

Framing a discrete choice experiment around a well-defined hypothetical scenario helps participants understand the survey and improves the accuracy of their choices. After being introduced to the topic, each participant was asked to imagine that summer was ahead and a preventive app that supports sun protection, helps with skin cancer prevention and promotes overall skin health was being recommended. The app would monitor daily duration and intensity of sun exposure and provide targeted feedback on when and how to protect. The 2 app options were provided in an unlabeled manner (as App 1 and App 2) in order to entirely steer the focus of our participants on the attributes and their levels [23].

The discrete choice experiment was generated in German and proofread by a native speaker. The reviewer was asked to pay attention on word choice, grammar, and punctuation as well as overall understanding.

Piloting

We piloted the initial discrete choice experiment face to face with 8 participants (5 males and 3 females), recruited in the waiting room of the University of Zurich Travel Clinic. Participants received the full questionnaire and were asked to share their thoughts while completing it. Through this think-aloud process, we captured how the discrete choice experiment's wording and layout could be improved to facilitate its feasibility. All conversations were recorded. Time to completion was measured to ensure that the survey could be

answered within a 20-minute time frame. The limit was set at 20 minutes to be below the average waiting time in the travel clinic's waiting room, where final recruitment occurred. Participants were also asked to provide feedback on the survey's relevance and ease of answering. Prior to that, pilot participants provided written informed consent and received written as well as verbal information on the study's aims and methodology. All participants considered the discrete choice experiment as interesting and relevant. Most commented on unclear terminology and expressed confusion of how the survey has to be answered. This led to (1) adjustments in the survey's wording, by choosing simple and self-explanatory terms, (2) changes in the survey's layout to reduce confusions of how choices should be made (eg, using colors to differentiate the 2 app options), and (3) revisions of the survey's instructions to ensure that participants understand how the survey has to be completed.

Survey Administration

The survey was administered at the University of Zurich Travel Clinic. On arrival at the clinic's waiting room, participants were informed about the option to participate, both verbally and through a flyer. Interested and eligible participants were asked to fill out the paper-based discrete choice experiment in the waiting room, where a member of the team was present to clarify questions. Eligibility required a minimum of 18 years of age, no current chronic conditions, a good understanding of German, and smartphone ownership. As scientific consensus on sample

size estimation of discrete choice experiments is lacking, we used a commonly applied rule of thumb, proposed by Johnson and Orme [24]. Thus, based on the number of choice tasks, alternatives, and analysis cells, we set our sample size at 200 participants. The paper-based discrete choice experiment's first pages aimed to inform participants, providing concise and easy-to-understand information on the study's aims, the topic of sun protection, each of the included attributes (with pictographs), and how to complete the discrete choice experiment.

Upon completion, participants were thanked for their time and effort with a sunscreen product. We also collected participant characteristics, including gender, age, highest attained education, perceived health, and perceived health app usefulness. We did not collect any personal identification or sensitive health information. Eligibility had to be confirmed by ticking the box "I confirm that I fulfill all eligibility criteria listed above" before answering the survey.

Analysis of Choice Data

Individual characteristics were explored descriptively. Participant preferences were estimated using conditional logit modeling, as proposed by McFadden [25]. The conditional logit regression is in line with random utility theory, according to which the utility a respondent $n=1,2,\ldots,200$ derives from an alternative j consist of a systematic (representative) and a random component. The systematic component is denoted by

where β_{Data}, $\beta_{Privacy}$, $\beta_{Sharing}$, $\beta_{Reminder}$, and β_{Cost} are the unknown parameters associated with the 5 attribute variables. $Data_{jn}$ takes the value 0 or 1 corresponding to manual entry or automatic entry, respectively; $Privacy_{jn}$ takes the value of 0 or 1 if consent for sharing data with commercial third parties is only given once or if consent is given on every sharing occasion, respectively; $Sharing_{jn}$ takes the value of 0 or 1 if data are not shared with a chosen health care provider or data are shared with a chosen health care provider, respectively; $Reminder_{jn}$ takes the value of 0 or 1 if noncustomizable or manually customizable, respectively; and $Cost_{jn}$ takes the value of 0 or 1 if the app has a one-time cost of 3 Swiss francs or if the app is

free, respectively. The results are reported as odds ratios (ORs), indicating relative importance of attribute levels compared to reference categories and their associated 95% confidence intervals. Although our survey entailed a binary cost option (free, 3 Swiss francs), we additionally ran the model with cost treated as a continuous variable, allowing it to express shifts in preference weights per 1 Swiss Franc change.

As a person's characteristics may influence their choice, we decided to individually test all collected individual characteristics for interactions with our attributes. These were identified and tested by repeatedly adding single interaction terms into the model. Multiple comparisons were counteracted through Bonferroni corrections ($\alpha=.002$). Significant interaction terms are reported individually. A mixed multinomial regression model was performed to explore potential preference heterogeneity. Here, we only assumed the effect of price to be random, as it is the only attribute where the assumption of an underlying latent continuous distribution is possible.

We additionally checked whether the 2 identical choice sets were answered consistently. In line with the axiom of completeness, participants who provide consistent answers are expected to choose the same alternative twice [17,26]. As directly excluding participants is not always recommended, we calculated percentages of inconsistent responses, assessed their distribution across individual participant characteristics and performed a sensitivity analysis comparing the results of our analyses with and without inconsistent responses [17,23,27]. Due to the small number of missing observations (4/3200, 0.13%) we decided to not perform multiple imputations. All analyses were performed in R (version 3.6.1).

Results

Participant Characteristics

Participant characteristic are presented in Table 2. Our sample (N=200) was unintentionally gender-balanced and yielded 3196 observations. Most participants were between 18 and 39 years old (135/200, 67.5%); had a tertiary education level (141/200, 70.5%), such as university and professional education degrees; rated their overall health as quite good (107/200, 53.5%) or very good (88/200, 44%); and had a positive perception (106/200, 53%) of apps' usefulness to their personal health.

Table 2. Participant characteristics.

Variable	Participants (N=200), n (%)
Gender	
Male	100 (50.0)
Female	100 (50.0)
Age (years)	
18-39	135 (67.5)
40-60	53 (26.5)
61-79	11(5.5)
80+	1 (0.5)
Highest education level	
Primary	5 (2.5)
Secondary	51 (25.5)
Tertiary	141 (70.5)
Other	3 (1.5)
Perceived health status (How would you rate your current overall health?)	
Very good	88 (44.0)
Quite good	107 (53.5)
Average	5 (2.5)
Quite bad	0 (0)
Bad	0 (0)
Perceived health app usefulness (Do you agree that apps could help you promote your health?)	
Fully agree	24 (12.0)
Agree	106 (53.0)
Uncertain	56 (28.0)
Disagree	10 (5.0)
Totally disagree	4 (2.0)

Main Effects

The results from the conditional logit model without interaction effects (model 1), are presented in Table 3 and in Figure 3. Overall, all 5 attributes contributed to consumer choices and overall preferences. The strongest driver of choice—denoted by the largest value of the OR—was the app's self-monitoring method, followed by cost, the option of data sharing with a health care provider, privacy control function, and reminder customizability (the least strong driver).

XSL•FO
RenderX

Table 3. Discrete choice experiment results from conditional logit regression (N=200).

Attribute and levels	OR[a] (95% CI)	P value
Self-monitoring method		<.001
Manual	Reference	
Automatic	2.37 (2.06-2.72)	
Privacy control		<.001
Single consent	Reference	
Multiple consents	1.57 (1.31-1.88)	
Data sharing with health care provider		<.001
No	Reference	
Yes	1.66 (1.40-1.97)	
Reminder customizability		<.01
No	Reference	
Yes	1.30 (1.09-1.54)	
Costs		<.001
Free	Reference	
3 Swiss francs	1.72 (1.49-1.99)	

[a]OR: odds ratio.

Figure 3. Forest plot from conditional logit regression (N=200). OR: odds ratio.
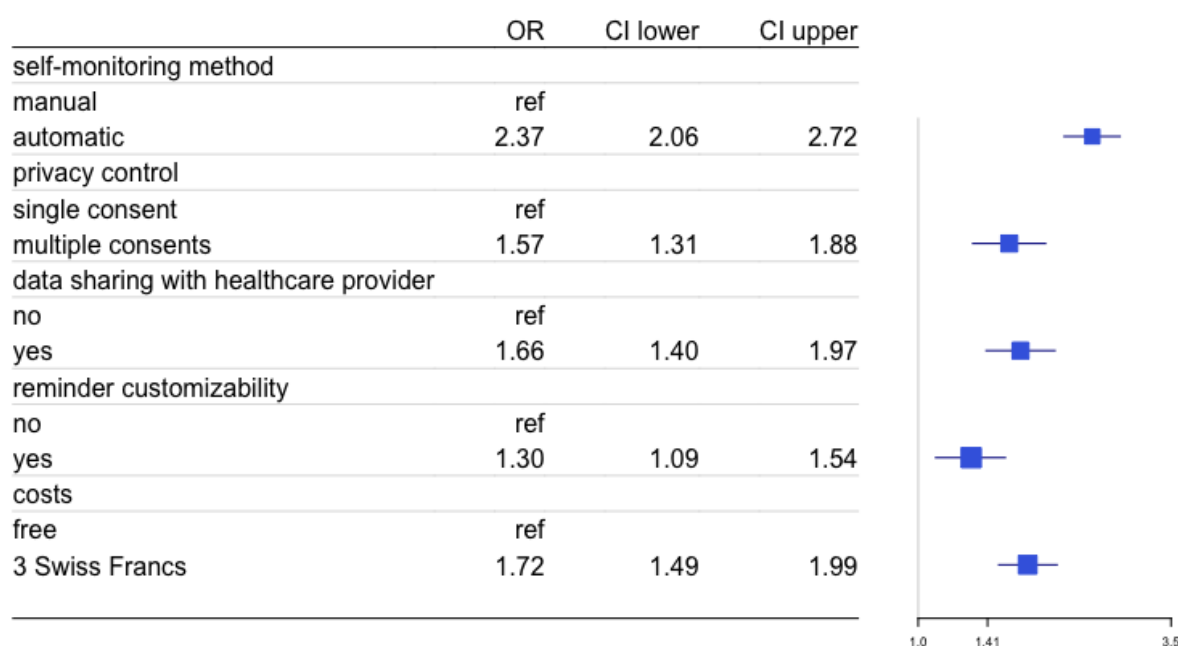
Our respondents strongly preferred automatic self-monitoring, with higher odds of choosing an app if it collected data automatically than if it required manual data entry (OR 2.37, 95% CI 2.06-2.72). Similarly, our respondents preferred no cost, with higher odds of choosing a free app than one requiring a one-time payment of 3 Swiss francs (OR 1.72, 95% CI 1.49-1.99). Modeling price as a continuous variable indicated that the odds of choosing an app alternative were reduced by 17% (OR 0.83, 95% CI 0.79-0.88) per 1 Swiss Franc increase in app cost. Respondents also preferred having the option of sharing their health data with a chosen health care provider than not having that sharing option (OR 1.66, 95% CI 1.40-1.97). Similarly favored was a thorough privacy control, with multiple user consents (given whenever app data are shared with third commercial parties) increasing the odds of choosing an alternative relative to one in which only a single consent was required (eg, after downloading the app) (OR 1.57, 95% CI 1.31-1.88). Finally, the setting of sun protection reminders was

the weakest, yet still a driver of choice, with customizable reminders being preferred over automatic reminders increasing the odds of choosing an app (OR 1.30, 95% CI 1.09-1.54). These findings were comparable to those of the mixed multinomial model (model 2), considering that as well as the unlabeled design and binary choice nature of our discrete choice experiment, we decided to limit reporting to the conditional logit's output and retain the mixed multinomial model as a sensitivity analysis.

Interaction Effects

After Bonferroni correction, only gender and perceived health app usefulness showed an interaction, in both cases with the self-monitoring method attribute. While both men and women preferred an app that generates their health data automatically, women had only half the odds of choosing an app alternative that offers that option (men: OR 3.42, 95% CI 2.78-4.21; women: OR 1.71, 95% CI 1.42-2.05). Those who were uncertain about whether apps can be helpful in promoting their health (neutral attitude) had less than half the odds of choosing an automatic self-monitoring app compared to those who fully agreed and thus had a positive perception of health app usefulness. Nonetheless, both groups preferred automatic generation over manual (full agreement: OR 4.19, 95% CI 2.71-6.48; neutral: OR 1.83, 95% CI 1.38-2.42). However, this effect changed direction for the group that fully disagreed with the statement that apps can be useful for personal health promotion (OR 0.32, 95% CI 0.11-0.92).

Inconsistent Responses

Eleven participants (11/200, 5.5%) provided inconsistent responses, that is, they answered the 2 identical questions differently; 6 were female and 5 were male. Most inconsistent responses were provided by participants between the age of 19 and 39 years (n=8), with a tertiary education (n=9), and who either had a rather positive or neutral perception (n=11) of apps' usefulness to their personal health. Comparing the results from the conditional logit model with and without the eleven inconsistent responders yielded comparable results; therefore, we decided not to exclude any participants.

Discussion

Principal Findings

This discrete choice experiment explored the preferences of healthy consumers for sun protection with a self-monitoring app. The survey included 5 attributes, presented as 2 unlabeled app alternatives (App A and App B) across 8 choice tasks (and 1 additional repeated task). We found that all 5 attributes influenced the choices of our participants. That confirms our preparatory qualitative work and the importance of these attributes for self-monitoring-based sun protection apps. Our hypotheses that health care consumers would prefer automatic over manual self-monitoring, customizable over noncustomizable reminders, and no cost over a single payment have all been confirmed. The method of self-monitoring had the strongest influence on app choice, followed by cost, data sharing with a health care provider, privacy control, and the customizability of reminders. Secondary analyses suggested

that the strength of preference for automatic self-monitoring was influenced by gender and perceived usefulness of health apps.

Effort and Ease of Use

Being well-aligned with current evidence, our participants mostly preferred sun protection apps that monitored health data automatically. Manual data collection requires time and effort, which is often perceived as tedious and inconvenient, reducing simplicity, ease-of-use and motivation [28,29]. Furthermore, the use of a sun protection app is more likely to occur during leisure time (eg, outdoor activities, holidays), which naturally makes manual monitoring a less attractive option. The underlying concept of effort and simplicity is recurring in the literature. For example, during the development of SunSmart, an ultraviolet monitoring and risk communication app, as well as mISkin, a sun protection app for those going on holidays, participants highly valued ease-of-use, which was described by the authors as a key feature [30,31]. This also holds for other prevention areas that require frequent and often complex monitoring, such as dietary intake and weight loss [32]. Our findings are potentially reinforced by the fact that self-monitoring sensors have become more accurate over the years, which in turn increases trust in automatic self-monitoring and facilitates trends toward easy-to-carry, and in fact, hidden monitoring devices [33,34].

Data Sharing

Although privacy concerns are often listed as key barriers of mobile self-monitoring apps, the attribute of privacy control, in the form of providing consents for data sharing with commercial parties, was relatively weak in influencing the app choices of our respondents [28]. Again, this may be explained (1) by the nature of sun protection behavior data and the good health status of our sample and (2) by the effort and time that is associated with the provision of multiple consents. Another factor that makes multiple consents less attractive is the information overload with which they are often associated [35]. Although the attribute did not drive choice as strongly as cost or the method of self-monitoring, our sample preferred providing multiple consents, repeated whenever their data are shared with commercial third parties. Our second data sharing attribute targeted the information flow between consumer and health care provider (eg, general practitioner, dermatologist). Having that sharing option clearly influenced app choice. On average, being able to share data with a doctor was more important than the app's privacy functions. This is a surprising and very interesting finding if we consider that our respondents were healthy and that such sharing functions are not yet widely available [28]. However, it potentially underlines the value and trust that app consumers place in interacting with health care professionals. Anderson et al [36] explored user experiences with self-care targeted mHealth apps and identified a similar pattern. Many users would have preferred the option of sharing their data with health care providers, allowing for a better-informed care and reduced visits [36]. In the context of sun protection, this preference might also be reinforced by the very nature of sun protection behavior data, which are not particularly controversial, stigmatized, or sensitive.

Reminders: Motivation Versus Nuisance

Equally unexpected was the relative importance of reminder customization, which was the weakest among the 5 attributes in influencing app choice. Reminders are a widely used mobile self-monitoring feature, highly applicable across all health areas. The insensitivity of reminders (eg, too frequent) has been recurrently cited as a barrier of mHealth use [28]. On average, our respondents preferred to be able to customize their app reminders, which may suggest previous experience with too many (or too few) reminders. The efficacy of reminders in health behavior change has been repeatedly demonstrated and so has their potential to be a nuisance factor and achieve the opposite [37-39]. Customizability is, therefore, often referred to as a key element of effective reminder systems. For example, during the iterative development of the mISkin sun protection app, prospective users expressed their preference to be able to personalize the frequency of reminder messages, as the default failed to meet their needs [31]. Beyond sun protection, a recently evaluated smoking cessation app identified that nonpersonalized reminders were primarily perceived as an irritation [37]. Similarly, reviews in the context of medication adherence and vaccinations described reminders, as well as reminder customizability, as being of key importance to user preferences and effectiveness [38,39].

Free First

On average, our respondents preferred free apps, which is an expected finding for 2 main reasons: First, healthy consumers are likely to use a sun protection app only periodically, such as during summer or sunny holidays. That potentially decreases an app's perceived value and the price someone is willing to pay for using it [40]. Second, our sample was predominantly young, which again might have influenced the overall willingness to pay. Gowin and colleagues [41] formatively explored the use of preventive and health promoting apps among college students, identifying that half of them would not be willing to pay, even if the amount was very small. This might be because of the high number of free apps, as well as because of different expectations by younger age groups, formed by their very early technology exposure [41]. Finally, younger age is associated with lower income, and thus, lower motivation to pay [42].

Limitations

Our findings should be considered with regard to the following limitations: First, recruitment occurred at a single center (Travel Clinic of the University of Zurich) and was subject to predefined inclusion criteria, for which reason our findings may not be generalizable. Second, despite the methodological robustness of discrete choice experiments, they maintain their feasibility by focusing on a limited number of attributes. Inherently, this reduces the discrete choice experiment's capacity to fully capture all possible influences of choice, whether individual or broader. We counteracted that by following a thorough qualitative attribute identification process and selecting attributes that are modifiable, are relevant, and cover an important array of mobile self-monitoring characteristics. Nonetheless, to fully understand these, we call for follow-up qualitative or mixed method research on mobile self-monitored sun protection and broad preventive mobile self-monitoring across population samples.

Conclusions

All 5 discrete choice experiment attributes influenced the choices of our respondents. Our sample's strongest preference was for sun protection apps that enabled automatic self-monitoring, which seemed to be even stronger for male respondents as well as those with an overall positive perception on the usefulness of health apps. Apps that were free of cost were the second most preferred choices. Furthermore, on average, an app was more likely to be chosen if it allowed for the in-app sharing of self-monitoring data with a chosen health care provider, for repeated user consents whenever data are shared with commercial third parties, and for reminders that can be customized. Surprisingly, the future-oriented function of data sharing with a health care provider was more important than the app's privacy control or reminder customizability. This underlines the potential value of connecting apps with trusted health care providers. Based on these findings, a preference- and user-sensitive self-monitoring app for sun protection should be simple and adjustable; require minimal effort, time or expenses; be interoperable; and have thorough as well as transparent privacy infrastructure. Factors that might partially explain our findings are our sample's good health and the rather periodic use of sun protection apps (eg, during summer). Similar features might be desirable for preventive health apps in other areas, paving the way for future discrete choice experiments.

XSL•FO

RenderX

Authors' Contributions

VN contributed to the conceptualization and implementation of the study, wrote the manuscript, and edited the manuscript. MM and MAP contributed to the conceptualization of the study, supervised the entire process, and edited the manuscript. JB contributed to the conceptualization of the study, provided statistical support, and edited the manuscript.

Conflicts of Interest

None declared.

References

1. Taylor K, Silver L. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center.: www.pewresearch.org; 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-11-08]

2. Measuring the Information Society Report 2015. International Telecommunication Union. Geneva; 2015. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2015.aspx [accessed 2019-11-18]

3. Poushter J. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies But advanced economies still have higher rates of technology use. Pew Research Center. 2016. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2019-11-18]

4. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. Research2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-11-12]

5. Kao CK, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]

6. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]

7. Roess A. The Promise, Growth, and Reality of Mobile Health - Another Data-free Zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]

8. About Melanoma. Melanoma Research Alliance. URL: https://www.curemelanoma.org/about-melanoma/ [accessed 2019-12-12]

9. Shapiro M, Johnston D, Wald J, Mon D. Patient-Generated Health Data - White Paper. RTI International. 2012. URL: https://www.itheum.com/resources/PGHD.pdf [accessed 2019-09-12]

10. Abril EP. Tracking Myself: Assessing the Contribution of Mobile Technologies for Self-Trackers of Weight, Diet, or Exercise. J Health Commun 2016 Jun;21(6):638-646. [doi: 10.1080/10810730.2016.1153756] [Medline: 27168426]

11. Murnane EL, Huffaker D, Kossinets G. Mobile health apps: Adoption, adherence, and abandonment. : Association for Computing Machinery; 2015 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers; September 7-11, 2015; Osaka. [doi: 10.1145/2800835.2800943]

12. Munson SA, Consolvo S. Exploring Goal-setting, Rewards, Self-monitoring,Sharing to Motivate Physical Activity. 2012 Presented at: 6th international conference on pervasive computing technologies for health care; May 21-24, 2012; San Diego p. 21-24. [doi: 10.4108/icst.pervasivehealth.2012.248691]

13. Buchwald A, Letner A, Urbach N, Von Entreß-Fürsteneck M. Insights into personal ICT use: Understanding continuance and discontinuance of wearable self-tracking devices. Insights into personal ICT use: Understanding continuancediscontinuance of wearable self-tracking devices. 26th ECIS; 2018 Presented at: 26th ECIS; June 23-28, 2018; Portsmouth p. 23-28.

14. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]

15. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ 2000 Jun 03;320(7248):1530-1533 [FREE Full text] [Medline: 10834905]

16. Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. Qual Health Care 2001 Sep;10 Suppl 1:i55-i60 [FREE Full text] [doi: 10.1136/qhc.0100055] [Medline: 11533440]

17. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]

18. Nittas V, Mütsch M, Puhan MA. Preferences for Sun Protection With a Self-Monitoring App: Protocol of a Discrete Choice Experiment Study. JMIR Res Protoc 2020 Feb 08;9(2):e16087 [FREE Full text] [doi: 10.2196/16087] [Medline: 32130187]

19. Nittas V, Mütsch M, Ehrler F, Puhan MA. Electronic patient-generated health data to facilitate prevention and health promotion: a scoping review protocol. BMJ Open 2018 Aug 10;8(8):e021245 [FREE Full text] [doi: 10.1136/bmjopen-2017-021245] [Medline: 30099392]

20. Hensher DA, Rose JM, Greene WH. Applied choice analysis: a primer. Cambridge: Cambridge University Press; 2005.

21. Aizaki H, Nishimura K. Design and Analysis of Choice Experiments Using R: A Brief Introduction. Agricultural Information Research 2008;17(2):86-94 [FREE Full text] [doi: 10.3173/air.17.86]

22. The R Project for Statistical Computing. R Foundation. Vienna URL: https://www.r-project.org [accessed 2022-11-08]

23. Walsh S, O'Shea E, Pierse T, Kennelly B, Keogh F, Doherty E. Public preferences for home care services for people with dementia: A discrete choice experiment on personhood. Soc Sci Med 2020 Jan;245:112675. [doi: 10.1016/j.socscimed.2019.112675] [Medline: 31760321]

24. Johnson R, Orme B. Getting the Most from CBC. Sequim: Sawtooth Software Research Paper Series. 2003. URL: http://www.sawtoothsoftware.com/download/techpap/cbcmost.pdf [accessed 2019-11-02]

25. McFadden D. Conditional logit analysis of qualitative behavior. Berkeley: Institute of Urban and Regional Development, University of California; 1973.

26. Rezaei A, Patterson Z. Detecting, NonTransitive, Inconsistent Responses in Discrete Choice Experiments. CIRRELT. 2015. URL: https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2015-30.pdf [accessed 2019-11-22]

27. Janssen EM, Hauber AB, Bridges JFP. Conducting a Discrete-Choice Experiment Study Following Recommendations for Good Research Practices: An Application for Eliciting Patient Preferences for Diabetes Treatments. Value Health 2018 Jan;21(1):59-68 [FREE Full text] [doi: 10.1016/j.jval.2017.07.001] [Medline: 29304942]

28. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]

29. Kim J. A qualitative analysis of user experiences with a self-tracker for activity, sleep, and diet. Interact J Med Res 2014 Mar 04;3(1):e8 [FREE Full text] [doi: 10.2196/ijmr.2878] [Medline: 24594898]

30. Dunstone K, Conway C. There is an app for that! Communicating UV via the SunSmart app. 2014 Presented at: NIWA UV Workshop; April 15-17, 2014; Auckland URL: https://pdfs.semanticscholar.org/bb37/8e7c335ea19ef0aa33d9d4fc7b884da99a9b.pdf

31. Rodrigues AM, Sniehotta FF, Birch-Machin MA, Olivier P, Araújo-Soares V. Systematic and Iterative Development of a Smartphone App to Promote Sun-Protection Among Holidaymakers: Design of a Prototype and Results of Usability and Acceptability Testing. JMIR Res Protoc 2017 Jun 12;6(6):e112 [FREE Full text] [doi: 10.2196/resprot.7172] [Medline: 28606892]

32. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Feb;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]

33. Pfeiffer J, Von Entress-Fürsteneck M, Urbach N, Buchwald A. Quantify-ME: Consumer acceptance of wearable self-tracking devices. 2016 Presented at: ECIS; June 12-15, 2016; Istanbul.

34. Xie W, Lewis WM, Kaser J, Ross Welch C, Li P, Nelson CA, et al. Design and Validation of a Biosensor Implantation Capsule Robot. J Biomech Eng 2017 Aug 01;139(8):1-10. [doi: 10.1115/1.4036607] [Medline: 28462423]

35. Banerjee S, Hemphill T, Longstreet P. Wearable devices and health care: Data sharing and privacy. The Information Society 2018;34(1):49-57 [FREE Full text] [doi: 10.1080/01972243.2017.1391912]

36. Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]

37. Struik LL, Bottorff JL, Baskerville NB, Oliffe JL. The Crush the Crave Quit Smoking App and Young Adult Smokers: Qualitative Case Study of Affordances. JMIR Mhealth Uhealth 2018 Jun 08;6(6):e134 [FREE Full text] [doi: 10.2196/mhealth.9489] [Medline: 29884602]

38. Stawarz K, Cox A, Blandford A. Don't forget your pill! Designing effective medication reminder apps that support users' daily routines. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26 - May 1, 2014; Toronto URL: https://doi.org/10.1145/2556288.2557079 [doi: 10.1145/2556288.2557079]

39. Pich J. Patient reminder and recall interventions to improve immunization rates: A Cochrane review summary. Int J Nurs Stud 2019 Mar;91:144-145. [doi: 10.1016/j.ijnurstu.2018.05.015] [Medline: 29958670]

40. Yuan S, Ma W, Kanthawala S, Peng W. Keep Using My Health Apps: Discover Users' Perception of Health and Fitness Apps with the UTAUT2 Model. Telemed J E Health 2015 Apr 28:144-145. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]

41. Gowin M, Cheney M, Gwin S, Franklin WT. HealthFitness App Use in College Students: A Qualitative Study. American Journal of Health Education 2015;46(4):223-230 [FREE Full text] [doi: 10.1080/19325037.2015.1044140]

42. Dinsmore JB, Swani K, Dugan RG. To "free" or not to "free": Trait predictors of mobile app purchasing tendencies. Psychology & Marketing 2017;34(2):227-244 [FREE Full text] [doi: 10.1002/mar.20985]

Abbreviations

CI: confidence interval
mHealth: mobile health
OR: odds ratio

XSL•FO
RenderX

Quantitative Methods for Analyzing Intimate Partner Violence in Microblogs: Observational Study

Christopher Michael Homan[1*], BA, MSc, PhD; J Nicolas Schrading[1*], BSc, MSc; Raymond W Ptucha[1*], BSc, MSc, PhD; Catherine Cerulli[2*], BA, JD, PhD; Cecilia Ovesdotter Alm[1*], BSc, MSc, PhD

[1]Rochester Institute of Technology, Rochester, NY, United States

[2]University of Rochester Medical Center, Rochester, NY, United States

[*]all authors contributed equally

Corresponding Author:
Christopher Michael Homan, BA, MSc, PhD
Rochester Institute of Technology
102 Lomb Memorial Dr
Rochester, NY, 14623-5608
United States
Phone: 1 5857196835
Email: cmh@cs.rit.edu

Abstract

Background: Social media is a rich, virtually untapped source of data on the dynamics of intimate partner violence, one that is both global in scale and intimate in detail.

Objective: The aim of this study is to use machine learning and other computational methods to analyze social media data for the reasons victims give for staying in or leaving abusive relationships.

Methods: Human annotation, part-of-speech tagging, and machine learning predictive models, including support vector machines, were used on a Twitter data set of 8767 #WhyIStayed and #WhyILeft tweets each.

Results: Our methods explored whether we can analyze micronarratives that include details about victims, abusers, and other stakeholders, the actions that constitute abuse, and how the stakeholders respond.

Conclusions: Our findings are consistent across various machine learning methods, which correspond to observations in the clinical literature, and affirm the relevance of natural language processing and machine learning for exploring issues of societal importance in social media.

Introduction

Intimate partner violence (IPV) encompasses physical violence, sexual violence, stalking, and psychological aggression (including coercive acts) by a current or former intimate partner [1]. The mental and physical consequences of IPV include depression, posttraumatic stress disorder, and suicidal thoughts and behaviors [2]. Physical consequences can include myriad acute and chronic health conditions, including but not limited to functional health status, cardiac health, complicated sleep histories, and higher reports of chronic pain [1,3-5]. These effects are long-lasting [1], and IPV affects people regardless of their sexual orientation or gender identity [6]. How IPV impacts any particular individual may depend on their childhood or adolescent experiences [7] or socioeconomic class [8]. For instance, women tend to be injured more severely and are killed more frequently than their male counterparts [8]. In the United States, 6.4% of men and 6.6% of women are affected by physical violence, sexual violence, or stalking annually [8]. Despite the similarity in frequencies between the genders, there are differences regarding severity and mortality, with more women reporting severe injury and dying as a result of IPV-related deaths.

A major gap in knowledge on the prevalence of IPV exists because population-level data are difficult to collect, particularly from victims [9]. Consequently, theories about why people become involved and remain in abusive relationships are based primarily on qualitative studies and surveys with small samples,

XSL•FO
RenderX

or larger samples of individuals who are often in the process of help-seeking. We know less about why people leave abusive partners because the process is often out of the vision of traditional service-providing agencies.

An alternate source of quantifiable data, such as Facebook, Twitter, or Instagram, is an alternate source of quantifiable data. It provides textual narratives at a level of personal detail reminiscent of focus groups or one-on-one interviews, but over populations larger than nearly any survey. Such an unsurpassed combination of scale and detail promises great rewards to the social, behavioral, and health sciences. Although these narrative tweets are short in length, they are a potentially rich source of information about IPV. Moreover, microblogging platforms such as Twitter offer a potential venue for public health prevention messaging that can be accessed broadly.

The amount of data present in social media is too great for human IPV experts to inspect manually. However, its semistructured, qualitative nature, along with the informality of the language used, means that human judgment is needed to make sense of the data. Machine learning has been used in many other settings to bridge this gap, and it can perform certain specialized tasks nearly as well as humans [10].

In this study, we present a machine learning workflow for social media data to hear, on a national, population-level scale, from victims themselves on why they stayed with abusive partners, or what helped them to leave. As a case study, we use this workflow to perform a comparative study on Twitter data tagged by #WhyIStayed or #WhyILeft. Such hashtags provide prelabeled data for machine learning.

Given the opportunistic nature of hashtagging, it is helpful to understand the history of these particular tags. On February 15, 2014 Ray Rice, a famous professional athlete in the National Football League, was arrested to assault his then fiancée (now wife), Janay Rice (née Palmer) [11]. First, news and sports media downplayed the incident. However, on February 19th, a video of the incident surfaced that showed Rice assaulting his fiancé in an elevator and eventually dragging her limp body out by her shoulders [11]. This caused an uproar in the advocacy and sports communities, and a backlash against the victim [11] (the backlash escalated after they married, one day after he was indicted by a grand jury). In response, on September 8, 2014 activist Beverly Gooden began using the hashtag #WhyIStayed to encourage victims of abuse to tell their stories about what kept them in abusive relationships; #WhyILeft appeared soon thereafter [12]. Tweets carrying these hashtags soon numbered in the tens of thousands, as people around the world decided to share their stories of IPV.

As an organizing principle and example of how to make sense out of the kind of opportunities for understanding and sensemaking that such movements provide, we focus on the specific question: *Why did both #WhyIStayed and #WhyILeft go viral?* Certainly, tweets disclosing stories about staying or leaving could be framed in terms of either hashtag, and adopting one or the other alone might have benefited the movement with a single (and thus easier-to-recall) catchphrase. These hashtags were invented by victims of IPV and virally adopted by a large community of victims over a relatively brief period of time. See Figure 1. Thus, they represent, among other things, victims of IPV not just telling their stories, *but doing so in their own terms,* rather than in those of health providers, researchers, or the criminal justice system. In a seminal work exploring 32 in-depth interviews with victims, Lempert [13] suggests that telling is a significant step in seeking help because it makes *public* their *fictions of intimacy* (citing [14]). Our primary goal was to assess how information gleaned through Twitter data could inform our understanding of survivors' lived experiences with staying in, or leaving, abusive relationships.

Figure 1. Counts per hour of #WhyIStayed (dotted) or #WhyILeft (solid) tweets from 9/8 to 9/12. Times in Eastern Standard Time, vertical lines mark 12-hour periods, with label corresponding to its left line. We removed spam from this set, but not meta tweets.
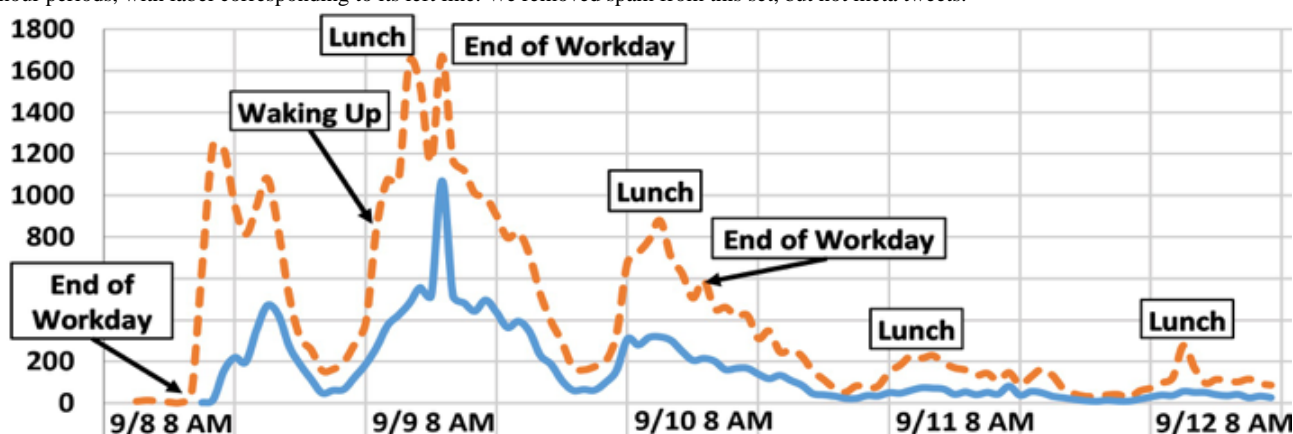
Methods

Data

We collected a corpus of tweets containing the hashtag #WhyIStayed or #WhyILeft using the Twitter and Topsy application programming interfaces, of which the latter is currently defunct. This corpus spans the beginning of September (the start of the trend) to the beginning of October 2014, when volume dropped to background levels. The majority of tweets are from the first week of the trend's creation (Figure 1).

Preprocessing

To partially anonymize the data, we replaced all URLs with the generic token "url." We removed spam tweets based on the usernames of prominent spammers and key spam hashtags such

as #MTVEMA, #AppleWatch, and #CMWorld. Additionally, we removed tweets in which Twitter accounts for DiGiorno Pizza tweeted *#WhyIStayed You had pizza*. Therefore, we excluded tweets containing tokens *pizza* or *digiorno*. The resulting corpus contained over 57,000 unique tweets.

Many tweets in the dataset were reflections on the trend itself or contained messages of support to those who shared their stories of abuse, for example, *not usually a fan of hashtag trends, but #WhyIStayed is incredibly powerful. #NFL #RayRice*. These instances, here denoted *meta-tweets*, were often retweeted, but they rarely contained reasons for staying or leaving (which were the interests of this study), so we excluded any tweets containing the keywords *janay/ray rice, football, tweets, trend,* and *video*.

Extracting Hashtag Labels

Typically, tweets disclosed reasons for staying (, respectively, leaving) and were prepended or appended with the hashtags #WhyIStayed (, respectively, #WhyILeft). *#WhyILeft because I gained the courage to love myself*. If the tweet contained only one of the target hashtags, the instance was labeled with that hashtag. For tweets marked by both hashtags, we split them into two identical instances, each distinctly labeled with one of the hashtags.

The resulting corpus comprised 24,861 #WhyIStayed and 8767 #WhyILeft labeled instances. This hashtag class imbalance may

be a consequence of the origins and media portrayals of the trend (the tweet that started the trend contained only the hashtag #WhyIStayed, and media reports tended to refer to the "#WhyIStayed" phenomenon) rather than an indicator that more victims stay than leave. The first #WhyILeft tweet occurred hours after the first #WhyIStayed tweet, and never gained as much traction (Figure 1).

To normalize comparisons between the tweets associated with each of these hashtags, we randomly sampled from the #WhyIStayed tweets to obtain a balanced set of 8767 examples per class. Of the 8767, we held out 1315 (14.99%) of this balanced set as a final *test set* for our machine learning experiments, and left the remaining 7452 (85.00%) as the *devset* (all of the remaining analysis in this section used the devset).

Manual inspection of the devset tweets revealed that, in addition to telling stories of IPV, the tweets served other purposes. To gain insight into the coarse grained language of these remaining tweets, we randomly sampled 1000 of them from the devset (473/1000, 47.30% #WhyIStayed and 527/1000, 52.70% #WhyILeft) and annotated them according to the coding scheme shown in Table 1, that is, as advertisements, jokes, tweets about leaving, meta (ie, tweets discussing or reporting on the #WhyIStayed/#WhyILeft phenomenon), tweets about staying, or *other*.

Table 1. Summary of labels from all four annotators, A1 through A4, compared with the gold standard. Each cell indicates the number of tweets an annotator gave a label to.

Annotator	Ads	Jokes	Leave	Meta	Other	Stay	Total
A1							
L[a]	3	6	356	57	67	38	527
S[b]	6	12	28	97	31	299	473
A2							
L	10	6	378	33	47	53	527
S	13	6	33	74	47	300	473
A3							
L	2	10	405	49	2	59	527
S	7	18	29	97	1	321	473
A4 [c]							
L	3	1	122	19	15	14	174
S	3	0	15	35	14	92	159

[a]L: #Why I Left.

[b]S: #Why I Stayed.

[c]A4 annotated only the first 333 tweets.

Before commencing annotation, to help better understand the distinct roles of each of the two hashtags played, we removed all occurrences of them from the tweets, to see if annotators could infer from the rest of the language whether the tweet was about staying or leaving without having the hashtag as a cue. We then studied the differences between tweets about staying versus leaving in terms of ngram bag-of-word models as

features, part-of-speech tags, and a restriction to subject-verb-object tweets only.

Machine Learning on Linguistic Features

One way to address our main research question is to ask whether #WhyIStayed and #WhyILeft hashtags indicated distinct classes of micronarratives or were merely framing devices. To address

this question, we trained naïve Bayes, linear support vector machine (SVM), and radial basis function (RBF) SVM classifiers from the scikit-learn python library [15], using the hashtags as ground truth, and various language features as input. We then use the SVMs to report the features that have the most predictive power. We also considered neural models [16,17], but it was harder to make sense of these results.

Results

Human Annotation

Four of the authors of this paper performed the annotation task. The overall agreement overlap was 0.77. Randolph's free-marginal multirater kappa [18] score was 0.72. We chose this multirater kappa because it allows any distribution of class labels that annotators assign (ie, it is free-marginal), unlike Fleiss' multirater kappa, which assumes a predetermined distribution [18]. According to the resulting annotations (Table 1), on average (over all annotators), 35.28% (1176/3333) of the instances were reasons for staying (S), 40.98% (1366/3333) were reasons for leaving (L), 13.83% (461/3333) were meta comments, 1.77% (59/3333) were jokes, 1.41% (47/3333) were ads, and 6.72% (224/3333) did not match prior categories (other).

The limited contextual information that such tweets provided sometimes made it difficult to interpret unambiguously. For instance *because i was slowly dying anyway* was marked as S by two annotators and L by the other two. The annotators disagreed on whether the victim decided to stay out of a sense of resignation or left because they felt there was nothing left to lose. (The ground truth label is #WhyILeft.) Another example of disagreement was *two years of bliss, followed by uncertainty and fear*. (This tweet's label is #WhyIStayed.) However, our results show that most tweets contained enough information for humans to infer their original hashtag labels, with annotators correctly identifying L more frequently than S.

Lexical Usage

Basic lexical statistics in the balanced devset *before* lowercasing, stoplisting, and lemmatizing are shown in Table 2. The top nine most frequent unigrams, bigrams, and trigrams of words in the balanced dataset *after* lowercasing, stoplisting, and lemmatizing are shown in Tables 3 and 4. In order for word and ngram counts to reflect the words representing *content* rather than the words serving grammatical *functions* or tweet *markup,* before making these counts, we stoplisted, lemmatized, and excluded *start-of-sentence* and *end-of-sentence* tokens from each tweet.

Each table reveals new insights into the #WhyIStayed/#WhyILeft movement, so we examine each in turn.

Table 2. Basic lexical statistics on the tokens and types in the two balanced sets. Types are unique tokens whereas hapax legomena are those tokens that only occur once in the data set.

Parameters	#WhyIStayed	#WhyILeft
Number of tokens	130,545	118,768
Number of types	7094	6269
Type:token ratio	0.054	0.053
Number of hapax legomena	3871	3340

Table 3. Top 9 most frequent unigrams (left) and bigrams (right) after preprocessing, with their respective frequencies in the Twitter devset.

Unigrams				Bigrams			
#WhyIStayed	Frequency, n	#WhyILeft	Frequency, n	#WhyIStayed	Frequency, n	#WhyILeft	Frequency, n
think	1061	love	930	think love	127	deserve better	298
love	971	realize	888	abusive relationship	112	finally realize	103
leave	872	want	702	feel like	95	realize deserve	80
abuse	754	leave	613	make feel	89	realize love	67
believe	578	know	594	try leave	78	want live	66
tell	550	better	570	emotional abuse	72	learn love	61
want	540	deserve	558	think deserve	67	want daughter	59
say	529	abuse	507	make believe	64	year old	56
know	518	life	497	kill leave	57	know deserve	55

Table 4. Top 9 most frequent trigrams after preprocessing, with their respective frequencies in the Twitter devset. The number in each cell indicates the number of times the trigram appeared in the dataset.

#WhyIStayed	Frequency, n	#WhyILeft	Frequency, n
make feel like	37	realize deserve better	56
pregnant hit url	25	know deserve better	40
stay abusive relationship	25	finally realize deserve	19
change conversation url	22	son deserve better	18
leave man yell	21	true love hurt	18
abusive relationship url	20	daughter deserve better	17
man yell url	20	want daughter think	15
say kill leave	20	want daughter grow	15
church support spousal	19	daughter grow think	15

Unigrams

The six words appearing in both columns of the unigram table (*love*, *leave*, *abuse*, *want*, *and know*) reveal common themes shared by both hashtags (Table 3). Our bi- and tri- gram analyses (presented below) reveal differences in how these common words are used. However, from the unigram perspective, the commonality of these words suggests, in the case of *leave* and *abuse*, that both #WhyIStayed and #WhyILeft tweets are often framed in terms of a leaving event, and that both often acknowledge the abuse that happened. This is perhaps not surprising, as the tags themselves suggest looking back on past abusive relationships. *Love* and *want* reveal that strong emotional forces are associated with both hashtags. *Know* indicates that knowledge (or lack thereof) is associated with both hashtags.

Of the words found in one list only, *think* and *believe* in the #Stayed list and *realize* in the #Left list suggest a transition between staying and leaving that involves learning the truth about false beliefs. The presence of *tell* and *say* in the #Stayed list suggests that an abuser coerced, deceived, or in some other way, emotionally manipulated the narrator. *Deserve* and *life* in the #Left list express positive sentiments.

Bigrams

Deserve better—the most common bigram by far—indicates that a change in one's sense of self dignity or fairness plays a major role in leaving. Both *realize* and *deserve* occur three times in the #Left list, suggesting that increased awareness, particularly from the perspective of justice or fairness, were important forces in helping narrators leave abusive relationships (Table 3).

Two words (*love* and *deserve*) appear in both lists. Each is paired with *think* in #Stayed and *realized* in #Left, suggesting a change in awareness of what love and fair treatment really are, respectively. That *deserve* also appears with *know* for #Left suggests determination.

The presence of *emotional abuse* in the #Stayed list confirms that, to the narrators of these tweets, emotional manipulation played a role in their staying in abusive relationships.

One telling bigram pairing is *kill leave* from the #Stayed list and *want live* from the #Left list. They indicate that people both remained and left abusive relationships out of fear for the safety of self and loved ones (the latter being supported by considerations about children, such as *year old*, *want daughter*, in the #Left list).

Trigrams

The most frequent trigrams in each list *make feel like* and *realize deserve better*, respectively, make an interesting pair. They seem to indicate the important role emotional manipulation plays in staying, and that leaving is precipitated by a realization or epiphany, perhaps sometimes ones that break the spell of emotional abuse (Table 4).

Hit, *yell*, and *kill* all appear in the #WhyIStayed list, suggesting that violence is an important force in keeping people in abusive relationships. Notably, *man* appears twice in this list. The trigram *church support spousal* shows the important role institutions and moral or religious values play.

Better, *deserve*, and *daughter* each appear four times in the #WhyILeft list. This shows that concerns about the welfare of dependents (note that *son* also appears on the list) or a desire for a better life (and that such a life is deserved) drive decisions to leave. The most common word in the #Stayed list, *url* (which appears four times), is harder to interpret; recall that, to protect privacy, we removed all URLs and replaced them with this token.

Subject-Verb-Object Structures

Fourteen percent of the dev- and test-set tweets have a subject-verb-object (SVO) structure, in which (a) the abuser is doing something to the victim, or (b) the victim is explaining something about the abuser or self. Such SVO structures represent the largest proportion of the total number of dependency structures in this data. Thus, we focus on its exploration in both corpus analytics and automated classification. These SVO structures provide insight into the abuser-victim relationship while maintaining sentence-level structures large enough to convey or indicate syntactic relationships, which tend to be more interpretable than isolated words.

XSL•FO

RenderX

We used the following conditional model to identify the most *indicative* verbs in terms of predicting hashtag class (#WhyIStayed vs #WhyILeft), among SVO tweets. Starting at the lemmatized predicate verb in each dependency parse, whenever the predicate verb followed an abuser subject word and preceded a victim object word, we added it to a distribution conditioned on class. The abuser subject words were *he, bf, boyfriend, father, dad, husband, brother,* and *man* for a male abuser, *she, gf, girlfriend, mother, mom, wife, sister,* and *woman* for a female abuser, and finally, *pastor, abuser, offender, ex, x, lover, church,* and *they* were used as neutral references. The victim object words were *me, sister, brother, child, kid, baby, friend, her, him, man,* and *woman.* These are denoted here as *abusers onto victim* structures. Analogous methods were used to extract structures in which the victim was the subject. We then determined the most indicative verb predicates from these conditional frequency distributions using the following formula for each such predicate:

where $count_{left}$ and $count_{stayed}$ are the number of times the verb appears in #WhyILeft or #WhyIStayed tweets, respectively. Table 5 shows those where the ratio is greater than 0.70 and, to avoid a bias toward lower frequency verbs, the total count exceeds a threshold of 0.5% of the total number of instances. In the case of *abuser onto victim*, the resulting frequency threshold was 11, and in *the victim as subject*, it was 68.

Table 5. The most indicative (in the direction of staying) verbs for abuser onto victim and victim as subject in the tweets having subject verb object structures. An exclamation point (!) before a verb indicates negation (eg, the phrase *he did not love me* would give the verb !love). Each cell indicates the weight of each subject verb object structure, as an support vector machine feature.

Abuser onto victim	Weight of SVO[a] structure	Victim as subject	Weight of SVO structure
convince	0.95	realize	0.98
need	0.94	think	0.91
isolate	0.94	!think	0.91
promise	0.92	find	−0.88
love	0.90	learn	−0.88
!love	−0.89	believe	0.86
!hit	0.89	!know	0.84
have	0.87	try	0.80
leave	−0.80	felt	0.73
tell	0.80	know	−0.71
be	0.78	tell	0.71
find	0.76	get	−0.70
choke	−0.75	N/A[b]	N/A
kill	−0.74	N/A	N/A

[a]SVO: subject-verb-object.

[b]N/A: not applicable.

Table 5 shows that physical *abusers on victim* verbs like *choke* and *kill* are associated with #WhyILeft, whereas for the *victim as subject* verbs *realize* appears as the most indicative verb in the data, along with *find* and *learn*. Additionally, a predominance of verbs about cognitive manipulation appear in #WhyIStayed tweets, such as *convince*, *promise*, *believe*, *think*, *!think* (where "*!*" denotes negation, for example, *I did not think he would...*) and *tell*. Heise et al [19] suggested that emotional dependence and an optimistic hope for change are reasons for staying, which these manipulative verbs seem to corroborate. Other interesting findings are the equal and opposite effects of *love* and *!love*, and the verb *!hit*, which suggests that perhaps the narrators believed they were not experiencing abuse because it was not physical, or that they feared physical retribution if they tried to leave.

Machine Learning on Linguistic Features

We used naïve Bayes, logistic regression, linear SVM, and RBF SVM classifier methods to automatically predict whether a tweet was tagged with #WhyIStayed or #WhyILeft. The RBF SVM method performed slightly better than the others, achieving a maximum accuracy of 78% (SD 1%) on the devset and 78% on the test set using a subset of features and hyperparameters: max df=12%, C=10, gamma=1. To better understand which linguistic features and preprocessing steps were most important to these classifiers, we performed feature ablation, following the procedure in Fraser et al [20], to determine the most important features the classifier used for prediction. Interestingly, the SVO features combined with ngrams worsened performance slightly, perhaps because of trigrams capturing the majority of SVO cases, but likely also because they just covered a small fraction of the dataset. The highest accuracy, nearly 78% on the test set, used a combination of ngrams and retweet

counts for features and informal register (tone) replacement in the preprocessing step.

We then used the *confidence* score of the linear SVM (defined as the distance from the classifier's separating hyperplane in the feature space of the model) on each feature, taken as a single input to the SVM, as an estimate of that feature's discriminativeness, or ability to distinguish between the hashtag class labels (Tables 5 and 6). This method can be seen as an alternative to the ngram count, which measures the predictive power of each ngram, rather than its frequency [21].

Table 6. Top 10 features, with their linear support vector machine weights using ngrams and retweet counts as features, and informal register replacement during preprocessing. Except for *try leave*, the top features were all unigrams.

#WhyIStayed	SVM[a] weight	#WhyILeft	SVM weight
think	3.0	realize	3.3
believe	1.6	finally	2.4
convince	1.6	tired	1.7
tell	1.5	realise	1.4
say	1.3	daughter	1.4
try leave	1.1	son	1.4
money	1.0	die	1.3
abuser	0.9	strong	1.3
feel	0.9	kill	1.2
young	0.9	anymore	1.2

[a]SVM: support vector machine.

The SVM picked up on many of the same reasons for leaving and staying as those shown in Tables 3 and 4, but also revealed new ones, including *tired*, *finally*, and *strong*, which appear on the #WhyILeft list (Table 6). These seem to suggest less an epiphany or triggering crisis and more a sense that the narrator was aware of and tolerated abuse for a long time until it became too much to bear.

For staying, language about cognitive and verbal manipulation was prominent (*think*, *believe*, *convince*, *tell*, *say*, and *feel*). Several new reasons also appeared: *try leave*, *money*, and *young*. The phrase *try leave* backs up claims in clinical literature that it is often difficult to gain external support to leave, and that victims of abuse frequently go through cycles of abuse that involve leaving and returning multiple times [19]. Financial distress is often a key factor for staying [19,22], so it is no surprise that *money* appears as a top feature for the SVM. The word *young* suggests that many were too young to leave or too naïve (due to their youth) to recognize that their relationship was abusive.

Subject-Verb-Object Structures

Restricting the dev- and test-sets to just those instances having an SVO structure, we trained the naïve Bayes, linear SVM, and RBF SVM. The linear SVM performed best, yielding 72% accuracy.

Table 7 shows the top SVO structures using the *confidence* score of the linear SVM on each data item. Some interesting structures not found in Table 5 appear here. For example, the #WhyILeft list reveals interventions from nonabusers (*sister tell me*). Taking a closer look at the supporting tweets, for example *because my sorority sisters and roommates told me nothing about how he treated me was okay*, suggests that these SVO structures refer to social support to which the victim has access. In the #WhyIStayed class, *church tell me* once again shows that religious institutions can play a role in keeping victims in abusive relationships. Several tweets indicated that their church condoned abuse as a means of avoiding embarrassment and divorce, for example *because the church told me that it was my responsibility as a godly wife to not embarrass him and just pray*.

Table 7. Top 10 subject-verb-object features for #WhyIStayed and #WhyILeft, with their support vector machine weights. An exclamation point (!) in front of a predicate verb indicates negation.

#WhyIStayed	SVM[a] weights	#WhyILeft	SVM weights
he hurt me	1.1	he tell him	1.3
they !remember him	1.1	he !protect me	1.2
he need me	1.1	he !tell me	1.0
he convince me	1.1	he lie me	1.0
she convince me	1.1	he stab me	1.0
he give child	1.0	he do kid	0.9
he remind me	1.0	sister tell me	0.89
he wear me	1.0	she have baby	0.89
he !abuse kid	1.0	he strangle me	0.78
church tell me	0.99	he attack me	0.77

[a]SVM: support vector machine.

Discussion

Principal Findings

Our analysis shows that the process of leaving often involves a better understanding of the reality of an abusive relationship, or coming to terms with a long anticipated but hazardous decision to leave. The words and phrases found in our results tend to be about the *pressures* of staying versus leaving or the *dynamics* involved in leaving (Figure 2). Buel [22] explains why women may choose to stay in abusive relationships, including fear of retaliation, lack of financial independence, concern for their children, emotional dependence, lack of support from friends and family, fear of divorce and the potential to lose custody of their children, and/or an optimistic hope through love that their abuser will change. Children are often

a factor in keeping victims in abusive relationships, and many victims will finally leave an abusive situation once their children have grown. The words and structures found in our results support many of these observations (eg, *church tell me*, *emotional abuse*, and *daughter deserve better*).

Heise et al [19] explain why victims of abuse leave and describe the dynamics of leaving. For instance, an increase in violence sometimes triggers a realization that their abuser will not change, that things are going to get worse, that the violence is going to affect their children, that they may be killed, etc. Support from friends, family, or society often allows those abused to leave. In any case, leaving is frequently a difficult process, involving cycles of denial, self-blame, and doubt. We found many of these same pressures and dynamics in our results (eg, *realize love*, *want live*, *sister tell me*).

Figure 2. A pictorial summary of our results, grouped according to the forces that keep people in abusive relationships or cause them to leave and the dynamics involved in leaving. In the dynamics section, gray arrows denote pairs of textual features that represent opposing pressures and appeared on opposite lists in the same table (1-, 2-, or 3-gram, subject verb object, and support vector machine classification features). SVM: support vector machine; SVO: subject verb object.
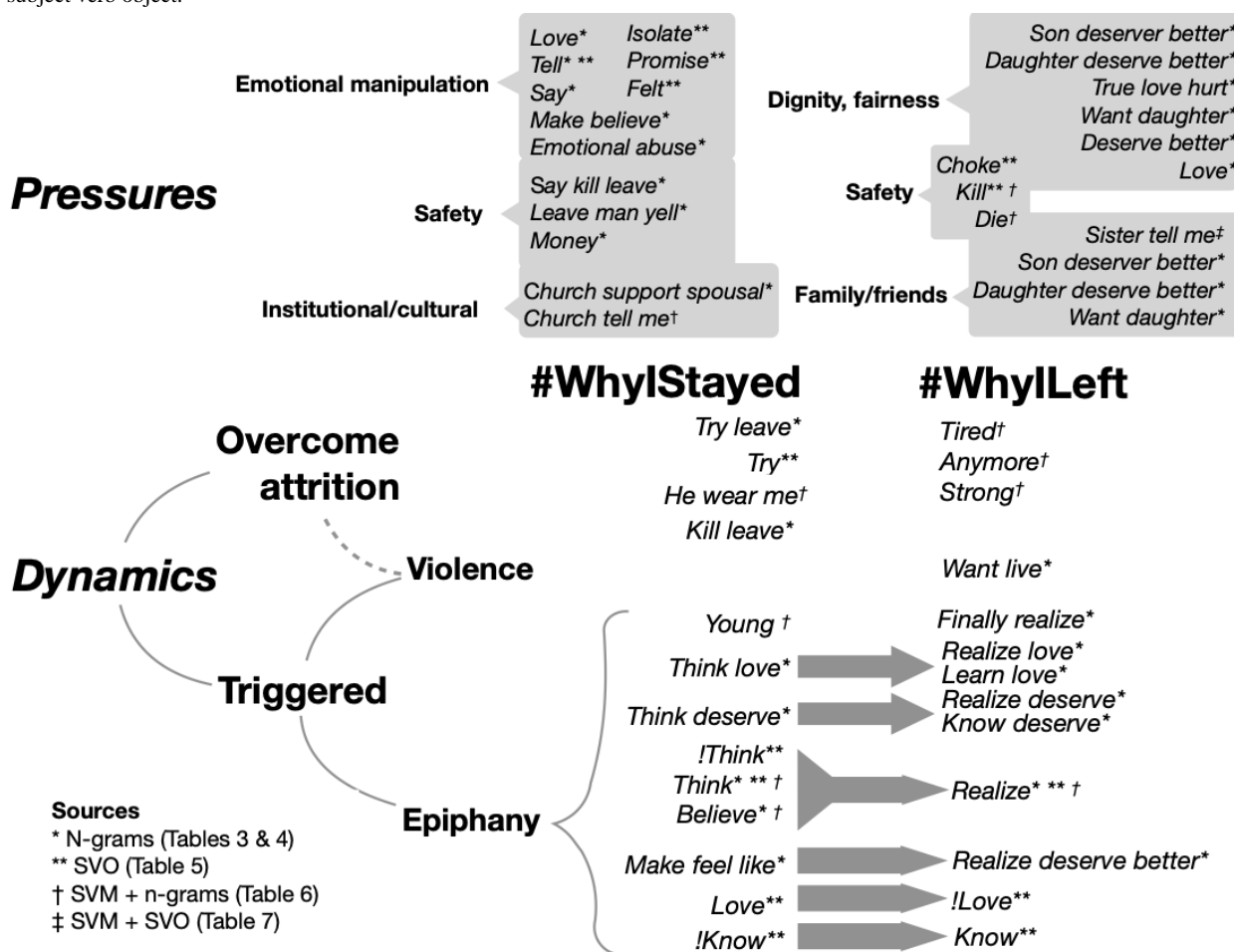
Pressures for Staying Versus Leaving

The main pressures we found for staying included emotional abuse, safety of self or other, and church communities. The pressures for leaving include a desire to be treated fairly or with dignity, the safety of self or other, and the concern for or support of a close family member.

The three pressures on each side interact with those on the other side in different ways. Emotional manipulation (a form of staying pressure) would seem to have less of an effect on victims as they become more aware of the injustice of their situation (leaving pressure). A sense of personal dignity (a leaving pressure) can cause victims to question the values of churches or other communities when they deny this dignity (a staying pressure).

For safety (both staying and leaving pressure), victims often say they stay in abusive relationships because they are afraid they will be physically or financially harmed if they leave, and they leave if they believe they or their children will be harmed or killed if they stay. Those who stay versus leave out of fear for personal safety are not necessarily the same individuals.

The last two pressures, church community and family or friends, are both about roles people outside the intimate partnership

play. There is a notable asymmetry between them, in that they represent different circles in a social ecology. Several researchers have noted disparities between how institutions and victims view IPV [23,24]. That church is the only community or institution specifically mentioned in our results is not too surprising, considering the importance of beliefs and knowledge on the dynamics of leaving a relationship and the role that churches play in preserving systems of beliefs. Lempert [13] discusses the pivotal roles that peer-level families and friends can play in helping victims leave. We found some evidence for this in our results, but the most frequent relationships mentioned were victims' children, who, as members of the family unit where the IPV occurred (and as frequent witnesses and victims themselves) were arguably in a different group from a peer-level family and friends, who were rarer in our results (Table 7).

Dynamics of Leaving

The dynamics of leaving can be roughly divided into two subgroups: those *triggered* by an event, such as an intervention by a friend, personal epiphany, or an imminent threat of harm, but before which the narrator was not expecting to leave, or even believed that their situation was unusually dire. The second group had to do with a slow wearing down, or *attrition,* where the victim was well aware of the harm the relationship was

causing, but had to weigh those harms against the costs of leaving.

Of these groups, words and phrases related to event triggering, specifically epiphanies, seemed to be the most common. Most of these were generic, such as *thought* or *realized*. However, victims did mention that, before leaving, they failed to understand the nature of love or violence, or did not realize that the abuse they experienced was not normal. Community pressures sometimes played a role in reinforcing the feeling that the abuse was normal. Sometimes the birth or maturation of children causes the narrator to see themselves more objectively. Sometimes they even witnessed the abuse they experienced visited upon their children, and this triggered a deeper awareness of the grim reality of their situation.

Dynamics around violence typically involve the narrator at some point becoming aware, either slowly over time or by a sudden escalation that they are at imminent risk of serious harm to death if they remain in the relationship. Heise et al [19] suggested that victims of abuse leave after an increase in violence triggers a realization.

In attrition dynamics, either the cost of abuse begins to outweigh the cost of leaving, or the victim, after some period of weakness or disempowerment, is able to summon the strength to leave. Heise et al [19] suggest that women are often *not* passive victims of abuse. Instead, they actively attempt to maximize the safety of themselves and their children, while struggling in secret to navigate the (often insufficient) support structures available to them. Many victims return repeatedly to their abusers before leaving permanently [19]. In our data, direct evidence of this dynamic was scarce in the ngram analysis, but abundant in the SVM analysis (Tables 6 and 7). However, the fear of personal harm from leaving due to violence, impoverishment, etc. was a significant pressure in many cases, evidence of which was abundant in the ngram analysis (Tables 3 and 4).

Ecological Model

Many of the findings from this study support a four-level ecological model [25,26] proposed by Heise et al [19] and expanded on by the World Health Organization [27]. All four levels that increase the likelihood that a man will abuse his partner are found in these data to varying degrees.

Individual

Ngrams like *hit* and *choke* (acceptance of violence as a means of solving issues), *childhood* (experiencing or witnessing abuse as a child)*,* and *want daughter, son deserve better* (trying to prevent their children from experiencing or witnessing abuse).

Relationship

Ngrams like *money* and *financial* (control of finances, economic stress) and the *abuser onto victim* verb *!love* (marital conflict)

Community

Ngrams like *try leave* and the *abuser onto victim* verb *isolate* (women's isolation), *church support spousal*, and *church tell me* (social groups that condone abuse).

Societal

Evidence for this last level was scarce; however, the SVO structure *he need me* suggests that abusers sometimes act out of frustration with societal norms or expectations.

Why Did Both #WhyIIStayed and #WhyILeft Go Viral?

Our results fall short of spelling out exactly why both the #WhyIStayed and #WhyILeft tags went viral, but they do paint a rich picture of the pressures and dynamics involved in staying versus leaving abusive relationships. Certainly, many reasonable explanations for the virality of these hashtags may have nothing to do with the actual stories of abuse disclosed. Perhaps in reaction to #WhyIStayed (which was tweeted first) activists adopted #WhyILeft to send a message that was more upbeat and empowering than #WhyIStayed. Certainly, the shift from emotional abuse to self-dignity and fairness comes with a shift from passivity to a more active and empowered role. One fairly clear, consistent pattern that we observed in our results was that leaving involved significant changes in life circumstances. Perhaps it is difficult even to recall the frame of mind one was in on either side of the leaving event without some kind of framing device like a hashtag, or perhaps having one hashtag for each side of this transition emphasizes the importance and significance of the transition itself in a way that having only one hashtag in the public sphere does not.

Limitations

We note several limitations in this research.

Bias Toward Female Victims and Male Abusers

We had hoped to study gender as a discriminative factor; however, instances with certain female abusers were rare (approximately 230 instances). Although it was difficult to determine a certain number, it appeared that the vast majority of the victims were female. This may be in part because males have significant inhibitions in reporting their abuse [8] and may therefore be less likely to tweet about their experiences and make their narratives public. It could also be that men do not face the same obstacles to leaving an abusive relationship, for instance, due to access to finances and/or alternative housing. Furthermore, we had no ability to stratify the data by sexual orientation, which could have implications for staying or leaving an abusive relationship for those that identify as lesbian, gay, transgender, or bisexual. There is a need for more research in this area.

Unique and/or Rare Forms of Abuse Missing

The properties of abuse and reasons for staying and leaving discovered in these data are affected by their relative frequency of occurrence. Unique and/or rare reasons for staying and leaving, and rare aspects of abusive relationships, may not be discovered using the methods presented here.

Noise

As with most social media data, it is important to know that these datasets likely contain posts by spam bots, lies by the users, or jokes that were missed by filters.

Handcrafted Pronouns and Lexical Items

The pronouns and lexical items used to convert the SVO features to *abuser onto victim* structures were handcrafted, potentially restricting the discriminative verbs that appear in the *Results* section.

Preprocessing

Lowercasing, stoplisting, and lemmatizing help to reduce dimensionality and thus improve learning performance, but case, tense, and certain ngrams that appear in the stoplist may be important features that were missed due to these preprocessing steps.

Length of Text

Although we recognize that these narratives are brief, qualitative methods such as free listing [28] often make use of brief texts. The notion of using words to create conceptual frameworks is not uncommon in mixed-method research. We suggest that our findings are worthy of continued, future exploration.

More broadly, our results show that when social media presents a large amount of data on a subject like IPV, even simple statistics such as ngrams can reveal a great deal of information about the nature of the subject. Machine learning methods, with their biases toward specific decision-making outcomes, reveal different insights. Although none of these approaches yielded effective predictive models, they provided data that were qualitative and quantitative enough to lend support to existing theories of IPV.

Comparison With Prior Work

Computational methods are applied to better understand the #WhyIStayed/#WhyILeft movement. Recent studies of this movement are concurrent with ours [29,30]. However, they were about their historical and qualitative aspects, or were based on a tiny sample of the data available. Our work is complementary to these studies, and our goal was to provide quantitative results that lend insight and credibility to these prior qualitative and clinical observations.

An extensive body of work explores how to extract affective information and other subjective signals from social media [31-35]. Adding part of speech tags to ngrams is often attempted as well as creating word classes via data inspection, using morphosyntactic features, and exploiting the sentiment of text instances. For instance, in Xu et al [36], linear models with ngrams are recommended for their simplicity and high accuracy, although in Lamb et al [37], word classes, Twitter-specific stylometry (retweet counts, hashtags, user mentions, and emoticons), and an indicator for phrases beginning with a verb were found to be helpful over ngrams on two different tasks.

Many of these works are motivated by commercial applications, for example, mining to extract individuals' sentiments about products or services [38,39]. Another stream of research focuses specifically on modeling, extracting, and/or tracking emotions on social media. Some of these works deal with emotions independent of context [40-42]. Other studies have studied their correlations with time [43,44] or other socioeconomic phenomena [45]. Still others model emotion as a social contagion [46-48] or focus on specific contexts, such as employment [49]. More recently, researchers have focused on specific emotional conditions or behavioral phenomena.

With respect to behavioral health, Coppersmith et al [50-52] built classifiers for detecting a number of mental health conditions, including major depression, posttraumatic stress disorder, seasonal affective disorder, and attention deficit hyperactivity disorder, by training tweets that match regular expressions related to each condition. De Choudhury et al [53,54] collected labels to consider as ground truth regarding the presence of major depression using crowd sourcing. Other researchers have focused on specific health issues, including posttraumatic stress disorder [51], early detection of epidemics [37,55], and bullying tweets [36,56]. A number of recent papers have studied suicidality on social media [57,58], or risk factors for suicide such as distress [59]. Closer to this paper and concurrent with our research, sexual abuse disclosures via anonymous Reddit were studied qualitatively by O'Neill [33] and quantitatively by Andalibi et al [60]. Subramani et al [61] performed a similar quantitative analysis of Facebook pages related to domestic violence. Karlekar and Bansal [10] studied the use of machine learning to extract narratives of sexual harassment from the SafeCity web-based forum.

Conclusions

The research presented here demonstrates the power of social media to uncover meaningful structural, semantic, linguistic, and textual characteristics, including actions, stakeholders, and situations related to abusive relationships. It revealed micronarratives in tweeted reasons for staying versus leaving abusive relationships. A classifier for distinguishing between tweeted reasons for staying versus leaving abusive relationships achieved an accuracy of 78%. Our textual analysis, in showing that partners leave violent relationships after an epiphany of self-realization, is validated in the clinical literature. Moreover, the sheer volume of data present in social media suggests the potential to learn more details about the nature and dynamics of interpersonal violence than—due to the stigma and shame related to disclosing stories of victimization that the #WhyIStayed/#WhyILeft movement has helped erode—are currently known and may potentially help clinicians to reduce the harm caused by abusive relationships.

There are a number of interesting directions for future work. For instance, social media data with course-grained geotags could be used to study whether reasons for staying and leaving differ across geographical locations, or how varying community-level characteristics of those locations (eg, poverty level, population density, education levels, etc.) affect IPV victims. Analysis of web abuse discourse across varied media would strengthen the present findings if they overlapped, and perhaps lead to a better understanding of how victims make sense of and manage IPV and abuse.

Conflicts of Interest

None declared.

References

1. Straus H, Cerulli C, McNutt LA, Rhodes KV, Conner KR, Kemball RS, et al. Intimate partner violence and functional health status: associations with severity, danger, and self-advocacy behaviors. J Womens Health (Larchmt) 2009 May;18(5):625-631 [FREE Full text] [doi: 10.1089/jwh.2007.0521] [Medline: 19445614]

2. Straus H, Cerulli C, McNutt LA, Rhodes KV, Conner KR, Kemball RS, et al. Intimate partner violence and functional health status: associations with severity, danger, and self-advocacy behaviors. J Womens Health (Larchmt) 2009 May;18(5):625-631 [FREE Full text] [doi: 10.1089/jwh.2007.0521] [Medline: 19445614]

3. Dichter ME, Cerulli C, Bossarte RM. Intimate partner violence victimization among women veterans and associated heart health risks. Womens Health Issues 2011;21(4 Suppl):S190-S194. [doi: 10.1016/j.whi.2011.04.008] [Medline: 21724140]

4. Pigeon WR, Cerulli C, Richards H, He H, Perlis M, Caine E. Sleep disturbances and their association with mental health among women exposed to intimate partner violence. J Womens Health (Larchmt) 2011 Dec;20(12):1923-1929 [FREE Full text] [doi: 10.1089/jwh.2011.2781] [Medline: 21988551]

5. Cerulli C, Poleshuck E, Raimondi C, Veale S, Chin N. "What fresh hell is this?" victims of intimate partner violence describe their experiences of abuse, pain, and depression. J Fam Violence 2012 Nov;27(8):773-781 [FREE Full text] [doi: 10.1007/s10896-012-9469-6] [Medline: 23226694]

6. Ard KL, Makadon HJ. Addressing intimate partner violence in lesbian, gay, bisexual, and transgender patients. J Gen Intern Med 2011 Aug;26(8):930-933 [FREE Full text] [doi: 10.1007/s11606-011-1697-6] [Medline: 21448753]

7. Whitfield C, Anda R, Dube S, Felitti V. Violent childhood experiences and the risk of intimate partner violence in adults. J Interpers Violence 2016 Jul 2;18(2):166-185 [FREE Full text] [doi: 10.1177/0886260502238733]

8. Smith SG, Chen J, Basile KC, Gilbert LK, Merrick MN, Patel N, et al. The National Intimate Partner and Sexual Violence Survey (NISVS): 2010-2012 State Report. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/violenceprevention/pdf/NISVS-StateReportBook.pdf [accessed 2020-10-17]

9. Burge SK, Becho J, Ferrer RL, Wood RC, Talamantes M, Katerndahl DA. Safely examining complex dynamics of intimate partner violence. Fam Syst Health 2014 Sep;32(3):259-270. [doi: 10.1037/fsh0000013] [Medline: 24911769]

10. Kalekar S, Bansal M. Understanding Diverse Forms of Sexual Harassment Personal Stories. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018 Presented at: EMNLP'18; October 31-November 4, 2018; Brussels, Belgium p. 2805-2811. [doi: 10.18653/v1/d18-1303]

11. Bien L. A Complete Timeline of the Ray Rice Assault Case. SBnation: Sports News, Video, Live Coverage, Community. 2014. URL: https://www.sbnation.com/nfl/2014/5/23/5744964/ray-rice-arrest-assault-statement-apology-ravens [accessed 2019-06-03] [WebCite Cache ID 78rLH91Yb]

12. Bahadur N. #WhyIStayed Stories Reveal Why Domestic Violence Survivors Can't 'Just Leave'. HuffPost India: Breaking News, Entertainment, Tech, Lifestyle. 2014. URL: https://www.huffpost.com/entry/whyistayed-twitter-domestic-violence_n_5790320 [accessed 2019-06-03] [WebCite Cache ID 78rKJ2ZVO]

13. Lempert L. The other side of help: negative effects in the help-seeking processes of abused women. Qual Sociol 1997;20(2):289-309. [doi: 10.1023/A:1024769920112]

14. Tifft L. Battering Of Women: The Failure Of Intervention And The Case For Prevention. Boulder, CO: Westview Press; 1993.

15. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn: machine learning in python. GetMobile: Mobile Comp and Comm 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]

16. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst 2017 Oct;28(10):2222-2232. [doi: 10.1109/TNNLS.2016.2582924] [Medline: 27411231]

17. Karpathy A. Char-RNN: Multi-Layer Recurrent Neural Networks (LSTM, GRU, RNN). GitHub: Where the World Builds Software. URL: https://github.com/karpathy/char-rnn [accessed 2019-06-03] [WebCite Cache ID 78rQNt2XR]

18. Warrens M. Inequalities between multi-rater kappas. Adv Data Anal Classif 2010 Oct 10;4(4):271-286 [FREE Full text] [doi: 10.1007/s11634-010-0073-4]

19. Heise L, Ellsberg M, Gottemoeller M. Ending violence against women. Popul Rep L 1999 Dec(11):1-43 [FREE Full text] [Medline: 11056940]

20. Fraser K, Hirst G, Graham N, Meltzer J, Black S, Rochon E. Comparison of Different Feature Sets for Identification of Variants in Progressive Aphasia. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology:

From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPsych'14; June 27, 2014; Baltimore, MD. [doi: 10.3115/v1/w14-3203]

21. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1-3):389-422 [FREE Full text] [doi: 10.1023/a:1012487302797]

22. Buel S. Fifty obstacles to leaving, aka, why abuse victims stay. Colo Law 1999;28(10):19-28 [FREE Full text]

23. Garcia-Moreno C, Watts C, Heise L. Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women. Department of Gender and Womens Health, World Health Organization. 2001. URL: https://www.who.int/gender/violence/womenfirtseng.pdf [accessed 2020-10-19]

24. Liang B, Goodman L, Tummala-Narra P, Weintraub S. A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence. Am J Community Psychol 2005 Sep;36(1-2):71-84. [doi: 10.1007/s10464-005-6233-6] [Medline: 16134045]

25. Bronfenbrenner U. The Ecology of Human Development. Cambridge, MA: Harvard University Press; 2009.

26. Heise LL. Violence against women: an integrated, ecological framework. Violence Against Women 1998 Jun;4(3):262-290. [doi: 10.1177/1077801298004003002] [Medline: 12296014]

27. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. Lancet 2002 Oct 5;360(9339):1083-1088. [doi: 10.1016/S0140-6736(02)11133-0] [Medline: 12384003]

28. Quinlan M. The freelisting method. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore; 2017.

29. Cravens J, Whiting J, Aamar R. Why I stayed/left: an analysis of voices of intimate partner violence on social media. Contemp Fam Ther 2015 Sep 25;37(4):372-385 [FREE Full text] [doi: 10.1007/s10591-015-9360-8]

30. Guidry J. Tweeting #whyileft and #whyistayed: Analyzing the Conversation Around Intimate Partner Violence on Twitter. In: Proceedings of the 143rd American Public Health Association Annual Meeting and Exposition. 2015 Presented at: APHA'15; October 31-November 4, 2015; Chicago, IL.

31. Pak A, Paroubek P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proceedings of the International Conference on Language Resources and Evaluation. 2010 Presented at: LREC'10; May 17-23, 2010; Valleta, Malta URL: https://www.researchgate.net/publication/220746311_Twitter_as_a_Corpus_for_Sentiment_Analysis_and_Opinion_Mining

32. Wang W, Chen L, Thirunarayan K, Sheth A. Harnessing Twitter 'Big Data' for Automatic Emotion Identification. In: International Conference on Privacy, Security, Risk and Trust and International Conferenece on Social Computing. 2012 Presented at: ASE/IEEE'12; September 3-5, 2012; Amsterdam, Netherlands p. 587-592. [doi: 10.1109/socialcom-passat.2012.119]

33. O'Neill T. 'Today I Speak': Exploring How Victim-Survivors Use Reddit. Int J Crime Justice Soc Democr 2018 Mar 1;7(1):44-59 [FREE Full text] [doi: 10.5204/ijcjsd.v7i1.402]

34. Kim S, Bak J, Oh A. Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations. In: Proceedings of the 6th AAAI International Conference on Weblogs and Social Media. 2012 Presented at: AAAI'12; June 4-7, 2012; Dublin, Ireland.

35. Hannak A, Anderson E, Barrett L, Lehmann S, Mislove A, Riedewald M. Tweetin' in the Rainxploring Societal-scale Effects of Weather on Mood. In: Proceedings of the 6th AAAI International Conference on Weblogs and Social Media. 2012 Presented at: AAAI'12; June 4-7, 2012; Dublin, Ireland.

36. Xu JM, Jun KS, Zhu X, Bellmore A. Learning From Bullying Traces in Social Media. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2012 Presented at: NAACL HLT'12; June 3-8, 2012; Montreal, Canada. [doi: 10.5555/2382029.2382139]

37. Lamb A, Paul M, Dredze M. Separating Fact From Fear: Tracking Flu Infections on Twitter. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: NAACl'13; June 9-14, 2013; Atlanta, GA.

38. Thelwall M, Buckley K, Paltoglou G. Sentiment in Twitter events. J Am Soc Inf Sci 2010 Dec 6;62(2):406-418 [FREE Full text] [doi: 10.1002/asi.21462]

39. Pfitzner R, Garas A, Schweitzer F. Emotional Divergence Influences Information Spreading in Twitter. In: Proceedings of the 6th AAAI International Conference on Weblogs and Social Media. 2012 Presented at: AAAI'12; June 4-7, 2012; Dublin, Ireland.

40. Mohammad S. #Emotional Tweets. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics. 2012 Presented at: SemEval'12; June 7-8, 2012; Montreal, Canada. [doi: 10.5555/2387636.2387676]

41. De Choudhury M, Gamon M, Counts S. Happy, nervous or surprised? Classification of human affective states in social media. In: Proceedings of the 6th AAAI International Conference on Weblogs and Social Media. 2012 Presented at: AAAI'12; June 4-7, 2012; Dublin, Ireland p. 435-438.

42. De Choudhury M, Counts S, Gamon M. Not All Moods re Created Equal! A Exploring Human Emotional States in Social Media. In: Proceedings of the 6th AAAI International Conference on Weblogs and Social Media. 2012 Presented at: AAAI'12; June 4-7, 2012; Dublin, Ireland.

43. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881. [doi: 10.1126/science.1202775] [Medline: 21960633]

XSL•FO
RenderX

44. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]

45. Bollen J, Pepe A, Mao H. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-economic Phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: AAAI'11; July 17-21, 2011; Barcelona, Spain p. 450-453.

46. Bollen J, Gonçalves B, Ruan G, Mao H. Happiness is assortative in online social networks. Artif Life 2011;17(3):237-251. [doi: 10.1162/artl_a_00034] [Medline: 21554117]

47. Kivran-Swaine F, Naaman M. Network Properties and Social Sharing of Emotions in Social Awareness Streams. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work. 2011 Presented at: ACM Conference on Computer Supported Cooperative Work and March; March 2011; Hangzhou, China p. 379-382. [doi: 10.1145/1958824.1958882]

48. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790. [doi: 10.1073/pnas.1320040111] [Medline: 24889601]

49. De Choudhury M, Counts S. Understanding Affect in the Workplace via Social Media. In: Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Media. 2013 Presented at: CSCW'13; February, 2013; San Antonio, Texas p. 303-316. [doi: 10.1145/2441776.2441812]

50. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology. 2014 Presented at: CLCP'14; June 27, 2014; Baltimore, Maryland p. 51-60. [doi: 10.3115/v1/w14-3207]

51. Coppersmith G, Harman C, Dredze M. Measuring Post Traumatic Stress Disorder in Twitter. In: Proceedings of the 8th AAAI International Conference on Weblogs and Social Mediaq. 2014 Presented at: ICWSM'14; June 1-4, 2014; Ann Arbor, Michigan, USA URL: https://doi.org/10.1038/s41598-017-12961-9 [doi: 10.1038/s41598-017-12961-9]

52. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter Through Self-reported Diagnoses. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology. 2015 Presented at: CLCP'15; June 5, 2015; Denver, CO p. -10. [doi: 10.3115/v1/w15-1201]

53. De Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th ACM International Conference on Web Science. 2013 Presented at: 5th ACM International Conference on Web Science; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]

54. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: Proceedings of the 7th AAAI International Conference on Weblogs and Social Media. 2013 Presented at: AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-137.

55. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. Artif Intell Med 2014 Jul;61(3):153-163. [doi: 10.1016/j.artmed.2014.01.002] [Medline: 24613716]

56. Xu JM, Burchfiel B, Zhu X, Bellmore A. An Examination of Regret in Bullying Tweets. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: NAACL HLT'13; June 9-13, 2013; Atlanta, GA p. 697-702.

57. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017 Aug;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001] [Medline: 29278258]

58. Luxton DD, June JD, Fairall JM. Social media and suicide: a public health perspective. Am J Public Health 2012 May;102 Suppl 2:S195-S200. [doi: 10.2105/AJPH.2011.300608] [Medline: 22401525]

59. Homan C, Johar R, Liu T, Lytle M, Silenzio V, Alm C. Analyzing Fine-grained Distress at Scale. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology. 2014 Presented at: CLPsych'14; June 27, 2014; Baltimore, MD p. 107-117. [doi: 10.3115/v1/w14-3213]

60. Andalibi N, Haimson O, de Choudhury M, Forte A. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, CA p. 3906-3918. [doi: 10.1145/2858036.2858096]

61. Subramani S, Vu HQ, Wang H. Intent Classification Using Feature Sets for Domestic Violence Discourse on Social Media. In: Proceedings of the 4th Asia-Pacific World Congress on Computer Science and Engineering. 2017 Presented at: APWC on CSE'17; Dec 11-13, 2017; Mana Island, Fiji p. 129-136 URL: https://arxiv.org/abs/1804.03497 [doi: 10.1109/apwconcse.2017.00030]

Abbreviations

IPV: intimate partner violence
RBF: radial basis function
SVM: support vector machine
SVO: subject-verb-object

XSL·FO
RenderX

Original Paper

The Relationship Between Images Posted by New Mothers on WeChat Moments and Postpartum Depression: Cohort Study

Weina Zhang[1,2], MSc; Lu Liu[1], BSc; Qijin Cheng[3], PhD; Yan Chen[2], BA; Dong Xu[4], PhD; Wenjie Gong[1,5,6], PhD

[1]XiangYa School of Public Health, Central South University, Changsha, China

[2]Changsha Hospital for Maternal and Child Health Care, Changsha, China

[3]Department of Social Work, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong, China

[4]Global Health and Health Systems, School of Health Management, Southern Medical University, Guangzhou, China

[5]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

[6]Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Wenjie Gong, PhD
XiangYa School of Public Health
Central South University
238 Shangmayuanling Lane, Xiangya Road
Kaifu District
Changsha, 410005
China
Phone: 86 13607445252
Email: gongwenjie@csu.edu.cn

Abstract

Background: As social media posts reflect users' emotions, WeChat Moments, the most popular social media platform in China, may offer a glimpse into postpartum depression in the population.

Objective: This study aimed to investigate the features of the images that mothers posted on WeChat Moments after childbirth and to explore the correlation between these features and the mothers' risk of postpartum depression.

Methods: We collected the data of 419 mothers after delivery, including their demographics, factors associated with postpartum depression, and images posted on WeChat Moments. Postpartum depression was measured using the Edinburgh Postnatal Depression Scale. Descriptive analyses were performed to assess the following: content of the images, presence of people, the people's facial expressions, and whether or not memes were posted on WeChat Moments. Logistic regression analyses were used to identify the image features associated with postpartum depression.

Results: Compared with pictures of other people, we found that pictures of their children comprised the majority (3909/6887, 56.8%) of the pictures posted by the mothers on WeChat Moments. Among the posts showing facial expressions or memes, more positive than negative emotions were expressed. Women who posted selfies during the postpartum period were more likely to have postpartum depression (P=.003; odds ratio 2.27, 95% CI 1.33-3.87).

Conclusions: The vast majority of mothers posted images conveying positive emotions during the postpartum period, but these images may have masked their depression. New mothers who have posted selfies may be at a higher risk of postpartum depression.

Trial Registration: International Clinical Trials Registry Platform ChiCTR-ROC-16009255; http://www.chictr.org.cn/showproj.aspx?proj=15699

KEYWORDS

social media; WeChat; WeChat Moments; postpartum depression

Introduction

In recent years, social media has become increasingly popular as a means for users to express their feelings and thoughts. Posts on social media platforms such as Facebook and Twitter can therefore serve as indicators or records of life events [1]. Worldwide, the number of monthly active users on Facebook and Twitter exceeds 2.6 billion and 330 million, respectively

[2,3]. In China, WeChat is the most popular social media platform, with more than 1.15 billion active monthly users [4]. Based on data from both the National Bureau of Statistics of China and Tencent's 2016 report [5,6], it is estimated that 72% of Chinese women (ages 15-80 years) use WeChat. Indeed, 60% of active users log into their WeChat accounts and post status updates with friends on WeChat Moments regularly [7]. Content posted on WeChat Moments is only visible to friends who follow one another. This social media platform is more private than Facebook and Twitter, for example [8]. Users are therefore more inclined to express their emotions authentically [8].

Some studies have used data collected from social media for sentiment analysis and have indicated that sentiment analysis has practical applications in many fields (eg, health care, finance, media, consumer markets, and government) and can facilitate the delivery of targeted information to people in said fields [9,10]. A review including 48 studies relevant to mental health issues used datasets from social media networks to research depression [11]. De Choudhury's [12,13] studies predicted the onset of postpartum depression and significant postpartum changes in mothers by analyzing shared Facebook and Twitter data. However, the review indicated that most related studies have focused on Western countries and have relied on text-based analyses [11]. The authors therefore suggested that future research should include more technical analyses, such as image-based analyses. Images posted on social media can be important information carriers as they contain large volumes of data. For example, photos posted on Instagram by depressed individuals are more likely to be bluer, grayer, and darker [14]. Furthermore, postpartum depression and maternal sensitivity have been associated with a lower proportion of photos of mothers' smiling babies [15], whereas selfies have been associated with a higher incidence of depression [16]. Certain facial expressions have been linked to depression and anxiety in social media users [17]. Theoretically, images posted on social media could be used to identify meaningful patterns in users' mental health. The initial basic function of WeChat Moments, the social media platform with the most traffic in China, is to upload images. However, due to the privacy settings of WeChat Moments, image data are difficult to obtain. To our knowledge, apart from our previous work in which we used WeChat Moments text emojis as features that were then modeled for sentiment analysis of perinatal

depression [18], no psychiatric study has used WeChat Moments to date.

In China, new mothers are expected to abide by a custom referred to as "doing the month," during which they stay home for about 45 days after childbirth and refrain from contacting people frequently [19]. This makes screening for postpartum depression particularly challenging. However, WeChat Moments posts offer a novel method to study the emotions of new mothers in China. Although postpartum depression screening currently relies heavily on self-reported questionnaires, some studies have found that people with depression are inclined to hide their symptoms of depression [20]. Moreover, self-reported screening methods are susceptible to social desirability response bias [21], which may cause false negative results during postpartum depression screening. Given the challenges inherent to the self-reported questionnaire approach, WeChat Moments posts could provide a new avenue for postpartum depression screening. Our cohort study was designed to analyze the features of the images that new mothers posted on WeChat Moments and to explore the correlation between these features and the mothers' relative risk of postpartum depression.

Methods

Participant Recruitment

Participants were recruited from two maternity and childcare centers in the cities of Changsha and Yiyang in Hunan province, China. We recruited women in the obstetrics clinics of these two centers from September 2016 to February 2017. The following inclusion criteria were applied: women who were pregnant, aged ≥18 years, and had a gestation period of ≤13 weeks (pregnancy weeks were estimated based on the first day of the last menstrual period).

The study was approved by the institutional review board of the Institute of Clinical Pharmacology of Central South University (ChiCTR-ROC-16009255). A total of 1126 women were recruited. In total, 15,647 WeChat Moments images and 6609 posts from 419 mothers, 123 of whom had postpartum depression, were used in this study. The participant recruitment and data collection process are depicted in Figure 1. The use of all the data from WeChat Moments in this study was authorized by the users.

Figure 1. Participant recruitment and data collection process flow diagram. EPDS: Edinburgh Postnatal Depression Scale. *Mothers did not complete the EPDS at 6 weeks postpartum, **Mothers refused to participate in the investigation, ***Mothers moved to another city for childbirth and were unable to continue participating in the study.
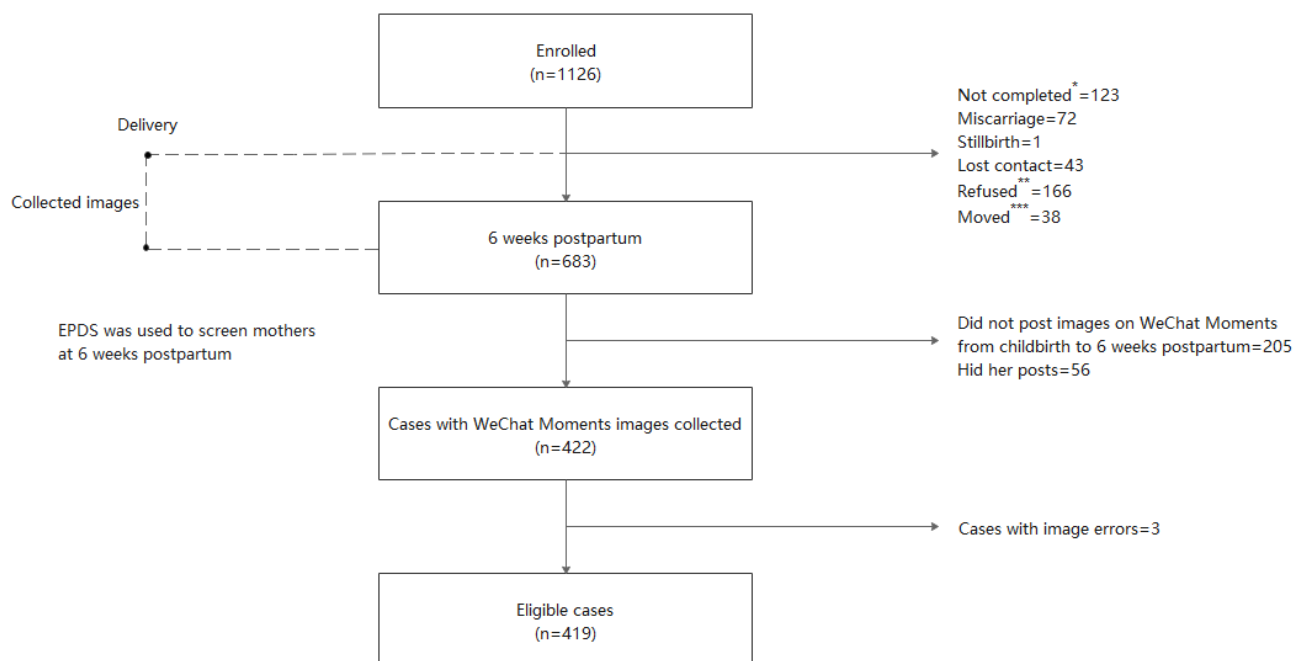
Measures

A number of different tools were used to collect data. We used questionnaires to measure the participants' perinatal depression levels and other psychological factors from the first trimester to 6weeks postpartum and collected the mothers' WeChat Moments data from childbirth to 6 weeks postpartum. The questionnaire was designed using simple items to investigate the participants' general demographic characteristics including age, first pregnancy, monthly income, education, and history of depression. A token gift of 10 RMB (US $1.5) was given to each participant who completed a questionnaire.

We used the Pittsburgh Sleep Quality Index [22] to measure the sleep quality of the participants during the postpartum period. In this index, sleep quality is divided into different grades based on the total score, where 0-5 indicates very good sleep quality, 6-10 indicates good sleep quality, 11-15 indicates average sleep quality, and 16-21 indicates poor sleep quality.

We used the 7-item Generalized Anxiety Disorder Scale [23] to measure the anxiety of the participants during the postpartum period. A total score of 10 or higher on this scale is indicative of anxiety.

We collected the images that mothers had posted on WeChat Moments from childbirth to 6 weeks postpartum using the WebCrawler document assembler. The privacy settings allowing image visibility to online friends in WeChat Moments include "all," "last six months," "month," and "three days" (friends will not be able to see the WeChat Moments earlier than the selected timeframe), so images were collected every 2 days.

The Edinburgh Postnatal Depression Scale (EPDS) [24,25] was used to screen for postpartum depression. The EPDS is a 10-item self-rated questionnaire. Each item is scored from 0 to 3, with a total score ranging from 0 to 30. The Chinese language EPDS used in this study was translated by Wang Yuqiong [26]. The critical value was 9.5.

Image Preprocessing

For image analysis, the following 3 errors were preprocessed and removed: (1) images that returned link error messages or manifested incompletely, (2) images with a screen resolution less than 800x600, and (3) black images. The remaining eligible images were converted to readable formats such as JPG and PNG.

Feature Extraction

Several different types of information were automatically extracted from the collected WeChat Moments data. Information categories included the following: post coloring, post volume and frequency, and the time posts were made.

Coloring

Colors can be expressed in various color spaces. In this study, we used the hue, saturation, and value color model. These 3 color properties are commonly used during image analysis [14,17]. Hue describes an image's coloring on the light spectrum, with the color type falling between 0° and 360°. Lower hue values indicate redder colors, and higher hue values indicate bluer colors (eg, 0 is red, 60 is yellow). Saturation refers to the vividness of an image and ranges from 0 to 1. Low saturation makes an image appear gray and faded (eg, 0 represents no color and is a shade of gray). Value refers to the

brightness of an image, which ranges from 0 to 1. Lower brightness scores indicate a darker image (eg, 0 represents black). The aforementioned colors of all the images were calculated in individual units. The mean number of pixels in each image was computed to determine the hue, saturation, and value. The colors of the images were extracted using MATLAB (version 9.7; MathWorks) and converted with OpenCV (version 4.0.0; Intel).

Volume and Frequency

The posting volume (each post contained 1-9 images) and the frequency of the WeChat Moments posts were calculated for each individual. Referencing Hicks' and Brown's [27] classification of the frequency of Facebook use, we defined the frequency of WeChat Moments posts as "high" (≥30 posts a month), "medium" (4-30 posts a month), and "low" (<4 posts a month).

Timing

Late-night posts were defined as posts made between 10 PM and 6 AM. Depending on whether the mother posted late at night or not, we defined this as "Yes" or "No," respectively.

Image Annotation

Assessing Agreement Among Researchers

To ensure the quality of our study, 11 independent researchers were recruited and uniformly trained to manually identify the content and emotions of the collected images. In line with Reece and Danforth's study [14], each photo was categorized by 3 different researchers. The researchers were not given any information about the mothers who provided the images. When the researchers had divergent opinions about the images, the following 3 rules were applied: (1) if 2 of the 3 researchers were in agreement, we would use the majority opinion as the correct result; (2) if all the researchers disagreed on a particular image, the principal investigators would discuss and reevaluate the image; (3) if the researchers had objections regarding the aforementioned reevaluation result, the image was classified as "other." In total, 4.7% (742/15,647) of the collected images were classified as "other."

Labelling Image Content

Images posted on WeChat Moments can contain diverse content. Using the criteria of previous studies [14,17], we labeled all images using a system with 11 types of tags: (1) people, (2) objects (eg, bed, clothes), (3) animals (eg, dog, cat), (4) landscapes (eg, lakeside, prairie), (5) vehicles (eg, automobile, bus), (6) plants (eg, flower, tree), (7) food (eg, cake, noodles), (8) buildings (eg, apartment, office building), (9) memes, (10) WeChat Moments advertisements, and (11) others. Each image could only be labeled with 1 tag. When a picture had multiple components (eg, a person hugging a pet, a landscape with some people), the following 2 rules were applied: (1) if an image contained people and their facial expressions were clear, we used "people" as the tag to define the content of the image; (2) If the image did not contain a person, we would consider what the mother who posted the image was trying to express and then decide on the tag accordingly. We also calculated the total number of tags and the percentage of each tag type.

Characterization of Pictures of People

We used 2 measures to characterize the pictures of people. The first measure was the number and proportion of selfies (ie, close-up facial images or images of the same person appearing multiple times) and pictures of the mothers' children. Studies have shown that posting selfies on social media can have adverse psychological effects on women [16,28]. Furthermore, a study by Schoppe-Sullivan et al [29] suggested that mothers with a high need for identity recognition were more likely to post pictures of their children on Facebook. Therefore, if mothers loaded selfies or pictures of their children, we placed them into a "yes" category. If they did not, we placed them into a "no" category. If mothers posted both selfies and pictures of their children, researchers assessed and categorized the images according to the subject or focus of the pictures. If researchers were in disagreement, images were processed according to the agreement assessment procedure we have described. The second measure was the number of positive (eg, joy and surprise) and negative (eg, anger, disgust, fear, and sadness) facial expressions of the persons in any given collected picture. Referencing the study by Dáu et al [15], which found that women who posted fewer images of smiling babies were more likely to exhibit maternal depressive symptoms, we categorized the facial expressions presented in each picture as positive or negative and then counted the proportion of each type of expression for each mother. Based on the proportion of positive facial expressions, we classified mothers as "No (0%);" "Yes, with a low proportion of positive facial expressions (0%-50%);" and "Yes, with a high proportion of positive facial expressions (>50%)." The first measure demonstrated the content of the pictures, while the second denoted the emotional expressions of the people in the pictures.

Assessment of Memes Posted to WeChat Moments

Through the expressions and words in the collected memes, we asked the researchers to categorize the mood of the memes as either positive or negative. We then calculated the number and percentage of positive and negative moods in the collected memes. However, we used only "Yes" and "No" classification criterion in the logistic regression model. Depending on whether or not mothers had posted memes, they were categorized as either a "Yes" or a "No." If memes were posted, they were assigned a "Yes," and if memes were not posted, they were assigned a "No."

Statistical Analysis

Different features of mothers with or without postpartum depression were analyzed using the chi-square test and *t* test. While controlling for the factors associated with postpartum depression, a binary logistic regression was conducted to explore the association between the image features and postpartum depression using the image features as the independent variables and with or without postpartum depression as the dependent variables. The statistical analyses were performed using SPSS (version 25; IBM). The classification of images in this study is based on previous studies [14-16,27-29].

Results

Demographic Characteristics of the Participants

We collected a total of 15,647 images and 6609 posts from 419 mothers on WeChat Moments. The demographic characteristics of the sample are shown in Figure 2. The average participant age was 27.6 years (SD 3.8 years). Participants predominantly had a monthly income of 2000-5000 yuan (US $282-705) and

had completed undergraduate or college education. The average hue, saturation, and value of the images was 33.9 (SD 10.6), 0.3 (SD 0.1), and 0.6 (SD 0.1), respectively. During the postpartum period, mothers posted an average of 0.89 (SD 5.07) images per day. The incidence of postpartum depression was 29.4% (123/419). Sample comparison analyses were conducted to compare the demographic features of mothers with postpartum depression and the mothers without postpartum depression (Figure 2).

Figure 2. Characteristics of women with or without postpartum depression. PPD: postpartum depression.
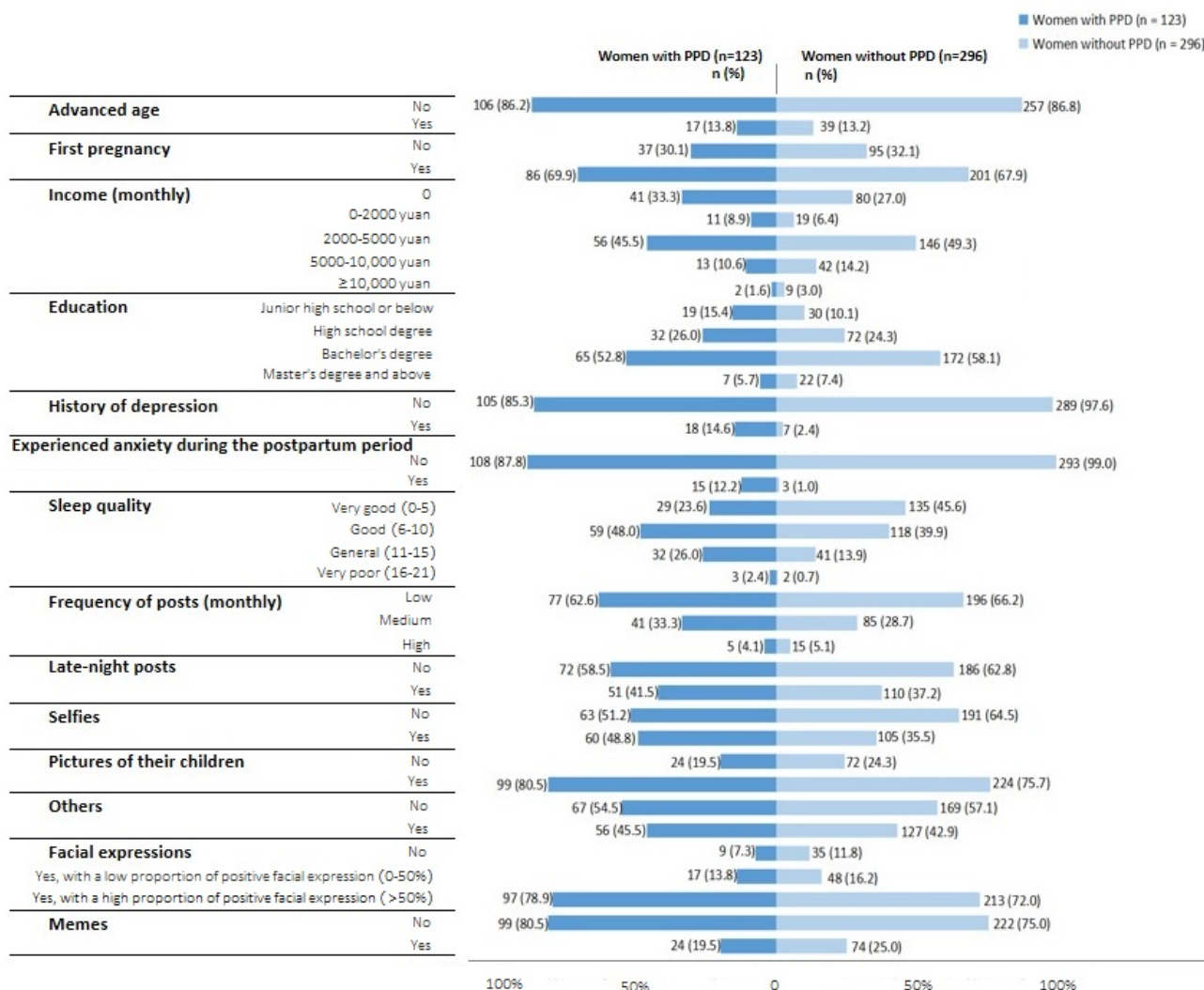
Image Analysis

Content

We calculated the number and proportion of each tag type (Figure 3A). Among the images posted by mothers during the

postpartum period, the highest proportion was pictures of people (6887/15,647, 44.0%), followed by pictures of objects (5356/15,647, 34.2%).

Figure 3. Content and emotions of the images. (A) The number and proportion of each tag type (n=15,647). (B) Content of pictures with people (n=6887). (C) The number and proportion of positive and negative facial expressions (n=6887). (D) The positive and negative moods in the memes posted by the mothers (n=375).
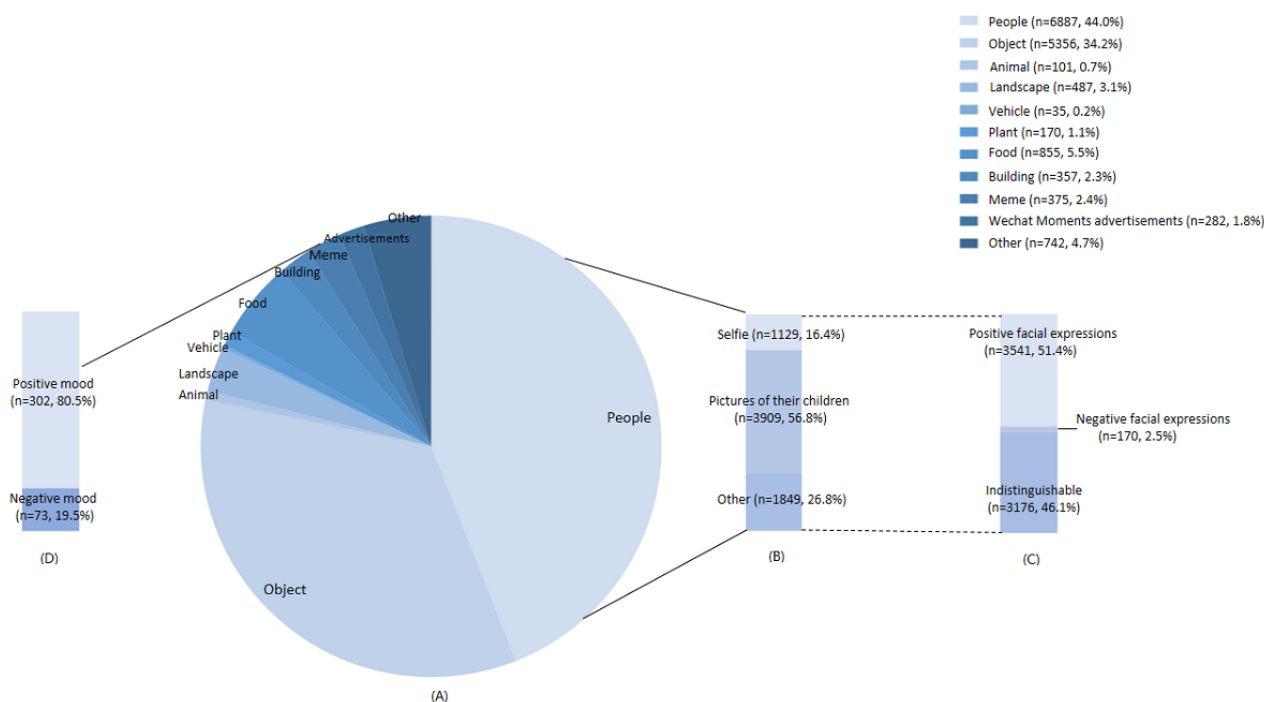
Pictures of People

Pictures of the mothers' children accounted for the highest proportion (3909/6887, 56.8%) of the pictures of people uploaded by the mothers during the postpartum period (Figure 3B), and most of the people in the pictures had positive facial expressions (3541/6887, 51.4%) (Figure 3C).

Memes

Mothers posted more memes depicting a positive mood (302/375, 80.5%) than a negative mood (19.5%) during the postpartum period (Figure 3D).

Correlation Between Image Features and Postpartum Depression

We used logistic regression to explore the relationship between the image features and the risk of 6-week postpartum depression (Table 1). In this model ($r^2 = 0.20$), selfies showed statistical significance after controlling for potential confounders. The factors associated with postpartum depression were included in the model as potential confounders. Women who posted selfies in the postpartum period were more likely to have postpartum depression (P=.003; odds ratio 2.27, 95% CI 1.33-3.87).

We put the variables of advanced age, first pregnancy, monthly income, education, depression history, anxiety, sleep quality, frequency of posts, late-night posts, colors, facial expressions, selfies, pictures of their children, others, and memes into the logistic regression model and presented the meaningful results in Table 1 (α=.05).

Table 1. Logistic regression analysis of the association between image features and postpartum depression risk.

Item	B	P value	Odds ratio (95% CI)
Anxiety	2.17	.002	8.71 (2.21-34.39)
Sleep quality			
Very good	—[a]	—	—
Good	0.90	.001	2.47 (1.42-4.28)
General	1.25	<.001	3.50 (1.74-7.01)
Very poor	1.91	.07	6.75 (0.84-54.55)
Selfies	0.82	.003	2.27 (1.33-3.87)

[a]Not available.

XSL•FO
RenderX

Discussion

Principal Findings

We found that, during the postpartum period, the mothers in this study tended to post pictures of their children and that the majority of the "people" in the collected pictures were categorized as experiencing positive emotions. However, the mothers who posted selfies during the postpartum period were more likely to have postpartum depression. Therefore, images posted by new mothers on WeChat Moments could be predictive of the mothers' mental health during the postpartum period.

Pictures of their children accounted for the largest proportion of images posted by the mothers (3909/6887, 56.8%), and 77.1% (323/419) of the mothers posted pictures of their children to WeChat Moments. This result is similar to that of the Bartholomew et al [30] study in which the proportion was 78.6%. Maternal identity confirmation may be the main reason that mothers share pictures of their children during the postpartum period. Johnson [31] suggested that social media was the main social platform mothers used to announce their pregnancies and the births of their babies. Another important reason mothers post pictures of their children may be that they need to strengthen ties with family and friends so that they can obtain more support after childbirth. The Gameiro et al [32] study found that mothers believed they received increased support from their nuclear family and friends in the pre- to postpartum period. Given that adults are more likely to give attention to babies' faces, many mothers who uploaded photos of their children to social media thought it was "likely" that the photos would be acknowledged (ie, commented on or "liked") by their social network friends [30]. It is possible that the majority of mothers consider posting pictures of their children as a particularly compelling way to engage with their online friends [30,33]. Accordingly, WeChat Moments may provide mothers with a platform through which to maintain strong ties with, for example, family and close friends. Meanwhile, sharing pictures of their children on WeChat Moments could be a particularly important means for mothers to identify themselves, build and maintain social capital, and find social support.

There were more positive emotions in the facial expressions of the people in the pictures and memes in this study than negative ones. In Chinese tradition, giving birth is a happy life event that should be celebrated by the whole family. Although we assumed that posting on WeChat Moments would reflect the feelings of the mothers in this study more accurately than posts made on other social media due to the WeChat Moment privacy settings, an element of social desirability response bias may have affected the expression of positive emotions. Pressures to portray one's best self may lead people to display deceptive versions of themselves on social media [28,34]. Social media users tend to present the happiest and most ideal sides of themselves even if these versions do not align with their actual emotions or experiences [35,36]. The findings of this study suggest that even if a new mother shows positive emotions on WeChat Moments, the risk of postpartum depression cannot be ruled out.

The posting of selfies on WeChat Moments seems to indicate the need to pay more attention to possible depressive symptoms in the users, as the women in our study who posted selfies during the postpartum period were more likely to have postpartum depression. This supports the findings of previous studies, which have shown a positive association between selfies and depression [16,28]. One potential explanation for the trend is that mothers posting selfies intend to show an idealized version of themselves, which may cause them to focus on perceived self-deficiencies, thus leading to depression [16,28]. In addition, a study found that women who posted selfies felt more anxious, less confident, and less physically attractive after posting, indicating that selfies may have harmful effects on mental health [16]. Posting selfies may be indicative of mothers' concerns regarding others' opinions of their images as well as their desire for more social attention [27,28]. Posting selfies may be one of the things that mothers can do during the "doing the month" period when their social contacts may be weaker. However, excessive concern for others' opinions, especially those of close friends and family, could negatively impact women's happiness and increase their risk of postpartum depression [37].

Limitations

This study had several limitations. First, the convenience sampling method used in this study limited the extrapolation of the results. Second, WeChat Moments allows users to upload pre-processed images, and some women were accustomed to using filters to embellish their images. Our inability to remove the effects of filters on the original images may have led to some bias. However, we believe that filter tones may have reflected some of the users' emotional states [14]. Third, WeChat has a feature to hide certain posts from certain people, so we cannot be confident that we collected all the mothers' images on WeChat Moments, even though we asked for all posts to be open to us when we requested informed consent. Fourth, we collected images posted before the EPDS measurement. However, the EPDS measurement reflects symptoms over the 7-day period before the EPDS questionnaire is administered. This means that we could not be certain about the order or cause-and-effect relationship between the images posted and the presence of postpartum depressive symptoms. Longitudinal research is needed to detect whether posting selfies during the postpartum period has a detrimental effect on mothers' mental health. Lastly, we analyzed the relationship between the mothers' posted images and postpartum depression but ignored the text accompanying the posts. The text could have helped us understand the content of the images better. We believe that a combined analysis of both text and images could provide more information, and this could be a direction for future research.

Conclusions

WeChat Moments is an important social media platform for Chinese mothers to share pictures of their newborns. Although mothers tend to post images featuring positive emotions, these pictures may actually mask depression. Particular attention should be given to new mothers who have posted selfies, as they seemed to be at higher risk for postpartum depression.

Acknowledgments

Authors' Contributions

As the first author, WZ developed the initial manuscript and performed the statistical analyses, while WZ and LL helped recruit the participants and collected the data. QC, YC, and DX contributed substantially to the revision and refinement of the final manuscript. WG conceived the study, guided the overall design, acquired the funding, and supervised the data analysis and the writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Saha K, Torous J, Ernala SK, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med 2019 Nov 25;9(6):1197-1207 [FREE Full text] [doi: 10.1093/tbm/ibz028] [Medline: 30834942]
2. Statista. Facebook: number of monthly active users worldwide 2008-2020. 2020. 2020 Nov 4. URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/#statisticContainer [accessed 2020-11-04]
3. Statista. Twitter: number of monetizable daily active users worldwide 2017-2020. 2020. URL: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/ [accessed 2019-08-14]
4. Network CII. Analysis of WeChat monthly active users, WeChat friends and future WeChat development trend of Chinese netizens in 2019. 2020 Feb 21. URL: http://www.chyxx.com/industry/202002/836124.html [accessed 2020-02-21]
5. National Bureau of Statistics of China. Tabulation on the 2010 population census of the people's republic of China. 2010 Jul 23. URL: http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm [accessed 2010-07-23]
6. Tencent Penguin Intelligence. Influence report of Wechat. 2016 Mar 21. URL: https://tech.qq.com/a/20160321/007049.htm#p=1 [accessed 2016-03-21]
7. Tencent Penguin Intelligence. 2017 WeChat Users and Research Report. 2017 Apr 24. URL: https://tech.qq.com/a/20170424/004233.htm#p=1 [accessed 2017-04-24]
8. Dexi L, Jiahong Q, Changxuan W, Xiping L, Minjuan Z, Haifeng G, et al. Feasibility of detecting depressive users using quasi-private social text. Journal of Chinese Information Processing 2018;32(9):93-102. [doi: 10.3969/j.issn.1003-0077.2018.09.014]
9. Kumar A, Jaiswal A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency Computat Pract Exper 2019 Jan 17;32(1):e5017-e5017. [doi: 10.1002/cpe.5107]
10. Xinyi G, Jinfeng L. A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. 2010 Oct 22 Presented at: Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS); 2019-10-22; Granada, Spain p. 472-477. [doi: 10.1109/snams.2019.8931720]
11. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. J Med Internet Res 2017 Dec 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
12. De Choudury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 Apr Presented at: Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems; 2013-4; Paris, France. [doi: 10.1145/2470654.2466447]
13. De Choudury M, Counts S, Horvitz E, Hoff A. Characterizing Predicting Postpartum Depression from Shared Facebook Data. 2014 Feb Presented at: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work 2014; 2014-2; Baltimore Maryland USA. [doi: 10.1145/2531602.2531675]
14. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1):15. [doi: 10.1140/epjds/s13688-017-0110-z]
15. Dáu ALBT, Callinan LS, Mayes LC, Smith MV. Postpartum depressive symptoms and maternal sensitivity: an exploration of possible social media-based measures. Arch Womens Ment Health 2017 Feb;20(1):221-224. [doi: 10.1007/s00737-016-0650-4] [Medline: 27416930]
16. Mills JS, Musto S, Williams L, Tiggemann M. Selfie? harm: Effects on mood and body image in young women. Body Image 2018 Dec;27:86-92 [FREE Full text] [doi: 10.1016/j.bodyim.2018.08.007] [Medline: 30149282]
17. Guntuku S, Preotiuc-Pietro D, Eichstaedt, JC?Ungar, LH. What Twitter profile and posted images reveal about depression and anxiety. 2019 Apr Presented at: Proceedings of the International AAAI Conference on Web and Social Media 2019; 2019-4-4; Munich, Germany.

18.   Chen Y, Zhou B, Zhang W, Gong W, Sun G. Sentiment Analysis Based on Deep Learning and Its Application in Screening for Perinatal Depression. 2018 Jun 18 Presented at: IEEE Third International Conference on Data Science in Cyberspace (DSC); 2018-6; Guangzhou,China. [doi: 10.1109/dsc.2018.00073]

19.   Pillsbury BL. "Doing the month": confinement and convalescence of Chinese women after childbirth. Soc Sci Med 1978 Jan;12(1B):11-22. [Medline: 565536]

20.   van der Steen Y, Myin-Germeys I, van Nierop M, Ten Have M, de Graaf R, van Dorsselaer S, et al. 'False-positive' self-reported psychotic experiences in the general population: an investigation of outcome, predictive factors and clinical relevance. Epidemiol Psychiatr Sci 2019 Oct;28(5):532-543 [FREE Full text] [doi: 10.1017/S2045796018000197] [Medline: 29656729]

21.   Poltavski D, Van Eck R, Winger AT, Honts C. Using a Polygraph System for Evaluation of the Social Desirability Response Bias in Self-Report Measures of Aggression. Appl Psychophysiol Biofeedback 2018 Dec;43(4):309-318. [doi: 10.1007/s10484-018-9414-4] [Medline: 30191358]

22.   Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

23.   Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

24.   Costafreda SG, Chu C, Ashburner J, Fu CHY. Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS One 2009 Jul 27;4(7):e6353 [FREE Full text] [doi: 10.1371/journal.pone.0006353] [Medline: 19633718]

25.   Meltzer-Brody S, Boschloo L, Jones I, Sullivan PF, Penninx BW. The EPDS-Lifetime: assessment of lifetime prevalence and risk factors for perinatal depression in a large cohort of depressed women. Arch Womens Ment Health 2013 Dec;16(6):465-473 [FREE Full text] [doi: 10.1007/s00737-013-0372-9] [Medline: 23904137]

26.   Wang Y, Guo X, Lau Y, Chan KS, Yin L, Chen J. Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale. Int J Nurs Stud 2009 Jun;46(6):813-823. [doi: 10.1016/j.ijnurstu.2009.01.010] [Medline: 19217107]

27.   Hicks S, Brown A. Higher Facebook use predicts greater body image dissatisfaction during pregnancy: The role of self-comparison. Midwifery 2016 Sep;40:132-140. [doi: 10.1016/j.midw.2016.06.018] [Medline: 27428109]

28.   Lamp SJ, Cugle A, Silverman AL, Thomas MT, Liss M, Erchull MJ. Picture Perfect: The Relationship between Selfie Behaviors, Self-Objectification, and Depressive Symptoms. Sex Roles 2019 Mar 5;81(11-12):704-712. [doi: 10.1007/s11199-019-01025-z]

29.   Schoppe-Sullivan SJ, Yavorsky JE, Bartholomew MK, Sullivan JM, Lee MA, Kamp Dush CM, et al. Doing Gender Online: New Mothers' Psychological Characteristics, Facebook Use, and Depressive Symptoms. Sex Roles 2017 Mar;76(5):276-289 [FREE Full text] [doi: 10.1007/s11199-016-0640-z] [Medline: 28239228]

30.   Bartholomew MK, Schoppe-Sullivan SJ, Glassman M, Kamp DCM, Sullivan JM. New Parents' Facebook Use at the Transition to Parenthood. Fam Relat 2012 Jul;61(3):455-469 [FREE Full text] [doi: 10.1111/j.1741-3729.2012.00708.x] [Medline: 23671354]

31.   Johnson S. "Maternal Devices", Social Media and the Self-Management of Pregnancy, Mothering and Child Health. Societies 2014 Jun 13;4(2):330-350. [doi: 10.3390/soc4020330]

32.   Gameiro S, Moura‑Ramos M, Canavarro MC, Soares I. Network support and parenting in mothers and fathers who conceived spontaneously or through assisted reproduction. Journal of Reproductive and Infant Psychology 2011 Apr;29(2):170-182. [doi: 10.1080/02646838.2011.553950]

33.   Joinson A. Looking at, looking up or keeping up with people?: Motives and Use of Facebook. 2008 Apr Presented at: Proceedings of the Conference on Human Factors in Computing Systems, CHI; 2008-4; Florence, Italy. [doi: 10.1145/1357054.1357213]

34.   Michikyan M, Dennis J, Subrahmanyam K. Can You Guess Who I Am? Real, Ideal, and False Self-Presentation on Facebook Among Emerging Adults. Emerging Adulthood 2014 Apr 24;3(1):55-64. [doi: 10.1177/2167696814532442]

35.   Jordan AH, Monin B, Dweck CS, Lovett BJ, John OP, Gross JJ. Misery has more company than people think: underestimating the prevalence of others' negative emotions. Pers Soc Psychol Bull 2011 Jan;37(1):120-135 [FREE Full text] [doi: 10.1177/0146167210390822] [Medline: 21177878]

36.   Kross E, Verduyn P, Demiralp E, Park J, Lee DS, Lin N, et al. Facebook use predicts declines in subjective well-being in young adults. PLoS One 2013;8(8):e69841 [FREE Full text] [doi: 10.1371/journal.pone.0069841] [Medline: 23967061]

37.   Smith AN, Fischer E, Yongjian C. How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter? Journal of Interactive Marketing 2012 May;26(2):102-113. [doi: 10.1016/j.intmar.2012.01.002]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale

XSL•FO
RenderX

Original Paper

Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach

Hannah Yao[1], MSc; Sina Rashidian[1], MSc; Xinyu Dong[1], MSc; Hongyi Duanmu[1], BSc; Richard N Rosenthal[2], MD; Fusheng Wang[1], PhD

[1]Stony Brook University, Stony Brook, NY, United States

[2]Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Fusheng Wang, PhD
Stony Brook University
2313D Computer Science
Stony Brook, NY, 11794
United States
Phone: 1 631 632 2594
Email: fusheng.wang@stonybrook.edu

Abstract

Background: In recent years, both suicide and overdose rates have been increasing. Many individuals who struggle with opioid use disorder are prone to suicidal ideation; this may often result in overdose. However, these fatal overdoses are difficult to classify as intentional or unintentional. Intentional overdose is difficult to detect, partially due to the lack of predictors and social stigmas that push individuals away from seeking help. These individuals may instead use web-based means to articulate their concerns.

Objective: This study aimed to extract posts of suicidality among opioid users on Reddit using machine learning methods. The performance of the models is derivative of the data purity, and the results will help us to better understand the rationale of these users, providing new insights into individuals who are part of the opioid epidemic.

Methods: Reddit posts between June 2017 and June 2018 were collected from *r/suicidewatch*, *r/depression*, a set of opioid-related subreddits, and a control subreddit set. We first classified suicidal versus nonsuicidal languages and then classified users with opioid usage versus those without opioid usage. Several traditional baselines and neural network (NN) text classifiers were trained using subreddit names as the labels and combinations of semantic inputs. We then attempted to extract out-of-sample data belonging to the intersection of suicide ideation and opioid abuse. Amazon Mechanical Turk was used to provide labels for the out-of-sample data.

Results: Classification results were at least 90% across all models for at least one combination of input; the best classifier was convolutional neural network, which obtained an F_1 score of 96.6%. When predicting out-of-sample data for posts containing both suicidal ideation and signs of opioid addiction, NN classifiers produced more false positives and traditional methods produced more false negatives, which is less desirable for predicting suicidal sentiments.

Conclusions: Opioid abuse is linked to the risk of unintentional overdose and suicide risk. Social media platforms such as Reddit contain metadata that can aid machine learning and provide information at a personal level that cannot be obtained elsewhere. We demonstrate that it is possible to use NNs as a tool to predict an out-of-sample target with a model built from data sets labeled by characteristics we wish to distinguish in the out-of-sample target.

(J Med Internet Res 2020;22(11):e15293) doi:10.2196/15293

KEYWORDS

opioid epidemic; opioid-related disorders; suicide; social media; machine learning; deep learning; natural language processing

XSL•FO
RenderX

Introduction

Background

The ongoing opioid crisis is characterized by an increasing number of deaths caused by opioid overdose; this number increased to such an extent that in 2017, the US Department of Health and Human Services declared a public health emergency [1]. In 2016, the US Centers for Disease Control and Prevention reported 42,000 deaths caused by opioid overdose, resulting from both prescribed and nonprescribed opioids and including intentional (suicide) and unintentional (accidental) deaths [2]. Since 2008, suicide has been the tenth leading cause of death in the United States. In 2016, suicide became the second leading cause of death for those aged 10 to 34 years [3]. The American Foundation for Suicide Prevention approximates 44,965 Americans die by suicide each year, and for each suicide, 25 people attempt suicide. These numbers may be underestimated because of the stigma associated with suicide [4].

Motivation

Individuals with chronic pain and poor mental health are often at risk of suicide; however, most current opioid overdose prevention methods neither assess suicide risk nor tailor prevention methods to personal situations [2,5,6]. Further hindering implementation of suicide screening is the lack of available data regarding death by overdose, frequently classified as *undetermined*, because of the inability to verify a death as a suicide. For example, if a deceased individual is found with a syringe by their side, it is difficult to determine whether it was an intentionally fatal dose or if the death was accidental (and without overdose intention) because of the risky route of administration by injection [6].

Individuals with opioid use disorder (OUD) are candidates for overdose due to the nature of the disorder. This study specifically focuses on suicidal ideation or intentions among individuals with a high possibility of OUD or, at minimum, opioid misuse. The Diagnostic and Statistical Manual of Mental Disorders defines OUD as a problematic pattern of opioid use leading to clinically significant impairment or distress [7]; approximately more than 2.5 million individuals are affected by OUD [8]. Diagnostic criteria include large amounts of opioids taken over a long period of time than intended, unsuccessful efforts to control opioid use, significant time investment in opioid-related activities or that necessary for recovery, impairment of daily life functionality, continued usage despite knowledge of consequences, dosage tolerance, and unavoidability of withdrawal [7].

The *New England Journal of Medicine* proposes that suicidal intention is blurred for individuals with OUD, who are repeatedly exposed to addiction-induced usage and withdrawal cycles and mood swings [2]. Ultimately, such prolonged exposure causes desensitization to risky and impulsive behaviors that can lead to death. Although the individual is aware of the consequences, the corresponding weight is muted and there is no associated conscious suicidal intent [9]. This aspect of OUD can lead to opioid overdose deaths but may not be explicitly considered in its contribution to overdose. Even in the absence of OUD, people who experience major depressive episodes, suicidal ideation, or other use and mental disorders are likely to misuse opioids, suggesting that opioid abuse and psychopathological conditions are self-perpetuating and cyclic behaviors [10].

Relevant predictors of suicide risk include chronic pain combined with depression [11,12]. Individuals with a history of depression also have a higher average daily dose and are more insistent on long-term opioid treatment for chronic pain [12], which is characteristic of OUD. Chronic pain and depression may also lead to suicidal thoughts even in originally healthy individuals [13], and both passive and active suicidal ideation are associated with mental illness, previous suicide attempts, and chronic pain [11,13]. Some suggest that opioid abuse manifests largely as a maladaptive coping strategy with physical and mental pain and is not solely biological [14,15]. Then, even if there is no conscious suicide intent, poor mental health leads the opioid-worn individual to adopt the mindset "I would not mind dying," and they can become more reckless in their behavior, resulting in a higher probability of overdose. Given these correlations, it is not unreasonable to seriously consider suicide risk among opioid users.

Those with OUD who experience thwarted belongingness may not be willing to discuss their concerns with those in real life due to social stigma [16]. Instead, they may turn to web-based means of discussion, connecting with and seeking support from those who are more likely to understand them than health professionals [17]. Most opioid studies based on social media have been conducted on Twitter. However, Twitter is limited to 280 characters, which may not provide enough space for struggling individuals to express themselves thoroughly, and it can be highly noisy due to lack of monitoring and the medley of topics [18-22].

In contrast, Reddit is a forum-based social media platform for the discussion of many topics. Each topic has its own *subreddit* for topic-specific content, and the character limit for a text post is 40,000. Users can interact directly with an original post and with other users by commenting below the original submission. According to Reddit's statistics page, it is currently the fifth most visited site in the United States, with more than 330 million monthly active users and 138,000 active communities [23]. Individuals are increasingly using mental health subreddits as a means of web-based support when offline support is unavailable. As Reddit also provides anonymity that is not available in real life, there is increased self-disclosure around stigmatic topics [24-26]. Furthermore, as subreddits are category-specific, struggling individuals can receive support and understanding from those with shared experience, who can provide therapeutic factors in the comments (eg, instillation of hope and altruism for those in opioid recovery or withdrawal) [27].

We consider the interpersonal psychological theory of suicide behavior as motivation for examining the connection between suicide and OUD [28]. In brief, this theory states that an individual will not die by suicide unless they have both the desire and the capability to act on that desire. According to the theory, suicidal desire is derived from an individual's feeling of burdensomeness and an impeded sense of belongingness

(social alienation) accompanied by the capability for suicide, which develops from repeated exposure to pain and exasperation [9]. Thwarted belongingness manifests as introspective torture in the individual with OUD, who often have ruined relationships or feels permanently labeled as *an addict* by society and cannot find understanding from those around them. The capability for suicide is acquired through active engagement with suicidal-like behavior and requires numbing out the innate sense of self-preservation [28]. This is applicable to those with OUD, many of whom have experienced overdose first hand or have witnessed their friends passing away from overdose and thus become desensitized over time.

Objectives

To the best of our knowledge, the intersection of suicidality and OUD has not been analyzed in a data-driven fashion with machine learning methods, although both areas have been examined separately and research has been conducted clinically. We seek to use subreddits in combination with each other and machine learning methods to classify for in-sample data then predict for out-of-sample data to detect suicide ideation in the context of OUD.

In this study, we (1) utilized the structure and nature of subreddits to train classifiers with the goal of extracting posts containing both OUD and suicidal ideation, (2) compared inputs and models for classification, and (3) aimed to extract suicide risk posts from an opiate context and extract opioid addiction from a suicidal context, where *context* refers to the subreddit(s) to which the data set belongs. We also asked nonspecialist workers from Amazon Mechanical Turk (MTURK) to annotate a sample of posts to obtain labels for whether a post is indicative of both suicide ideation and opioid addiction.

Related Work

Social media studies utilizing Reddit for its ability to provide massive amounts of text data have become popular. Many of these focus on current mental health concerns, such as depression and anxiety, and seek to analyze the linguistic features of post content and employ traditional methods such as N-grams, latent Dirichlet allocation topic modeling, linguistic inquiry and word count, term frequency-inverse document frequency (TF-IDF), and word embeddings to extract and analyze the emotional and mental states of Reddit users [25,27,29-31].

The subreddits r/suicidewatch and r/opiates revolve around the topics of suicidality (ie, discussions about, circumstance disclosure, emotional expressions, etc) and opioid usage (ie, drug doses, activities done while high, withdrawal anguish, etc), respectively. Analyses of specific subreddits have provided insights into the general mentality of subreddit users through their use of language. For example, a study by Park et al [29] showed that there are significant structure differences between sentences in mental health subreddits and nonmental health subreddits. They revealed differences between Reddit's r/suicidewatch and the general mental health subreddit in their readability indexes and showed increased usage of first-person singular pronouns and decreased usage of first-person plural, second-person, and third-person pronouns [29]. Kumar et al

[32] conducted a survey of r/suicidewatch to test for the Werther Effect, specifically following high-profile suicides, and found increased negative emotions, anger, self-harm, and suicidal words. They similarly reported increased usage of first-person singular pronouns along with decreased lexical diversity, longer posts, and fewer future-oriented words and more past-oriented words [32]. Yet another analysis of r/suicidewatch extracted recurring topics of discussion and then compared the results with suicide risk factors defined by domain experts; it was found that r/suicidewatch is capable of capturing less well-known dimensions of the risk factors, such as concerns about failing school, gun ownership, drug abuse, financial concerns, deceased friends, and domestic abuse [33]. The analysis of r/opiates yields dominant topics and characteristics of discussion, particularly for opioid users, such as opioid withdrawal and opioid intake routines [34].

These findings from r/suicidewatch support the findings of analyses of real-life writings left behind by people who attempted or completed suicide. In particular, a study of word use in poetry by suicidal and nonsuicidal poets revealed shifts in first-person singular pronoun and communication words (eg, *talk*), and a linguistic study of suicide completers showed that these individuals differ in time orientation compared to control groups [35,36]. The idea of *alienation* from the interpersonal theory of suicide supports the findings of increased usage of first-person pronouns—that when one is unable to relate to society, the presence of first-person plural *we* and other words of interaction decreases [28,37].

Neural networks (NNs) have been shown to achieve excellent results in natural language processing. Unlike traditional methods, *learning text from scratch* at the character or wording level requires no knowledge of sentiment; high-level targets can be used as input [38,39]. This is especially valuable in this context because drug-related words are often slang, and expressions related to getting high or intoxicated may be falsely categorized as negative when they are not due to the inability of sentiment lexicons to learn domain specificity [40]. Furthermore, there is no need for exhaustive feature engineering because weights can be learned [41].

Kim Yoon [42] demonstrated the ability of a simple convolutional neural network (CNN) with 1 layer of convolution in sentence classification for multiple data sets, finding that the performance of a simple CNN is comparable with that of traditional methods [42]. Orabi et al [43] used CNN and recurrent neural networks (RNNs) to predict depression for Twitter data, showing that CNN performs better than RNN. Johnson and Zhang [44] successfully used word sequences to classify documents with CNN, whereas Kim and Orabi only classified short, sentence length texts [44]. Singh et al [45] compared NNs and found that CNN performed best when considering both speed and accuracy. Therefore, we used a simple CNN architecture for its speed and performance.

Methods

Data Collection

The social media platform used in this study is Reddit. We focused on the subreddits r/suicidewatch and r/depression and opioid-related forums; also, a control group was used to model healthy, nondisordered language. The subreddit r/suicidewatch acts as a *bottom net*, such that if a user expresses suicidality, they will commonly be directed to r/suicidewatch for help. The subreddit has policies that forbid trolling and discourage activism (ie, repeatedly posting hotlines) in favor of direct peer support, which gives this forum credibility [46]. Subreddit r/depression is similar in forbidding *empty encouragement* (ie, *don't worry, it gets better*). These characteristics exemplify the seriousness of these discussion boards and substantiate the reliability of the data. In contrast to these subreddits, however, r/jokes has few or no rules, only specifying that there are to be no personal attacks and stating that discussions must be lighthearted and civil; meanwhile, r/showerthoughts simply asks for original content. Posts in opioid-related forums were assumed to be indicative of individuals who struggle with opioid abuse and seek understanding on the web; our assumption is based on the fact that the act of posting about opioids on social media indicates that the presence of these drugs is sufficiently dominant in one's life. All opioid-related forums were selected if the forum was relatively active and if the focus drug of the forum was frequently discussed in r/opiates, the most active opioid subreddit. Data were collected from Reddit using the pushshift.io Python Application Programming Interface (API) [47] and Reddit's Python Reddit API Wrapper (PRAW) [48]. The IDs of submissions between June 2017 and June 2018 were obtained via pushshift.io and then passed to PRAW, which retrieved the actual submission. A submission consists of an ID, author, title, and body text/content.

We used posts from r/suicidewatch to represent suicidal language and a collection of opioid-related subreddits to model the language of opioid users (which is referred to in this study as r/opiates or *opiates data*). We assumed that all posts from r/suicidewatch were suicidal and that the language of r/depression would be most similar to that of r/suicidewatch [49]. Several control subreddits were used for language comparison based on a study by Shen et al [50] for detecting anxiety on Reddit, which implies that subreddits themselves can be used as labels. This set of diverse subreddits seeks to offset the impact of excessive first-person pronoun usage that is common when expressing negative emotions and subject-specific words [51,52]. We assumed that the majority of posts in these control subreddits were *mentally healthy*, meaning that post content (saying nothing about the user themselves) is not reflective of mental disturbances such as depression and suicidal ideation. In addition to the selected subreddits from those used by Shen et al [50], we included r/careerguidance and r/personalfinance to account for finance stressors that may be experienced by those in r/suicidewatch, and we included r/offmychest to account for the potentially rougher and more profane language of opioid users, which is indicative of strong negative emotions but unlikely to indicate suicidality. Examples of posts in some specific subreddits are shown in Table 1, and the data groups and their subreddits are listed as follows:

- *Depression*: depression.
- *Suicidewatch*: suicidewatch.
- *Control*: askdocs, askscience, books, careerguidance, fitness, frugal, jokes, lifeprotips, offmychest, parenting, personalfinance, productivity, randomkindess, relationships, showerthoughts, talesfrom retail, theoryofreddit, wholesometextposts, writing, youshouldknow.
- *Opiates*: benzodiazepines, benzorecovery, heroin, methadone, opiates, opiatesrecovery, quittingkratom, suboxone.

Table 1. Example posts from subreddits belonging to data groups.

Subreddit group	Title	Body text
Control (r/showerthoughts)	It becomes less and less acceptable to cry in public the older you get, despite the reasons for doing so becoming more and more valid	Kids just don't understand
Opioid related (r/opiates)	Happiness is…a big, dark shot of Heroin after being sick all week. <3	And also knowing you have enough for not only the next morning but at least a few more days while you figure out what to do next!
Depression	I won't commit suicide but I wouldn't mind dying	So much shit has been piling on and on. I feel like I am not making the people I care about proud and the only reason they talk to me is because of pity. I will not take my own life, but if a car hit me, I got terminal illness or if something fell on me. I would not be sad about me being gone
SuicideWatch	I'm waiting for the courage to end my own life	I feel like I'm close to making a serious attempt sooner than later and I'm ok with that. My impulsive behaviors have gotten worse other the last few months and some of those ways include bodily harm. In September i impulsively jumped out of a friends window and injured myself and now I am cutting myself at random for the first time in years. No one around me understands how exhausting it is to wake up everyday in my own skin. In my own head. I'm sick of the stomach pangs and guilt and crying and disappointment. Some nights I just pray I'll have the courage to end it. People die in the world all over, shouldn't matter when I go

Feature Matrix Construction

We maintained the high-level structure of the language, only lowercasing and removing URLs [42]. The text was not lemmatized, and all stop words and profanity were kept. All punctuation marks were removed except for periods, commas, exclamation marks, question marks, and apostrophes. We sought to make use of word order for text categorization [44]. As the title of a post may be a good representation of post content, the title of every post was prepended to the body text. The joined title and body content composed the NN input. All posts were either zero padded or truncated to a length of 1500 words, as we wanted to use the entire post if possible. Each post (title+body text) was preprocessed, tokenized, and encoded, preserving the word sequence, and is thus represented as a vector of words. We also reran the classification using only posts with more than 1000 words to show that differing content lengths do not impede performance. Textbox 1 shows examples at each text-processing stage.

Textbox 1. Example of text-processing stages.

Unprocessed text

- Question. What's the best otc med to OD on? I'm so over the day to day grind. I'm a failure at life and ready to check out...

Processed text

- Question. what's the best otc med to od on ? i'm so over the day to day grind . i'm a failure at life and ready to check out . . .

Tokenized

- | question | . | what's | the | best | otc | med | to | od | on | ? | i'm | so | over | the | day | to | day | grind | . | i'm | a | failure | at | life | and | ready | to | check | out | . | . | . | . |

As sequential text data cannot be used as is for training machine learning models, there are several techniques that can convert these data to numerical values:

1. TF-IDF: The term frequency-inverse document frequency shows how important a word is in a document with respect to its frequency in other documents. We used a 100-dimension TF-IDF from in-sample unigrams and bigrams [53].
2. Word embedding: This is a set of methods and language models for converting text data to vector space. We took advantage of 3 methods, including the Gensim Word2Vec model Global Vectors for Word Representations (GloVe), which require embeddings to be trained in advance on a large corpora of data and also a simpler version of word to ID mapping, which was performed in running time [54,55].
3. Character embedding: As out-of-dictionary words are all mapped to random vectors in pretrained word embeddings or the slightest change in slang language could end up mapping words to different vectors, adding character embedding as well showed improvements for training the machine learning models.

In this study, we took advantage of different combinations of these knowledge representation methods to evaluate whether classification performance improves additional information by alleviating vocabulary issues. Multimedia Appendix 1 shows the complete list of combinations.

Models

Our main focus was on a CNN text classifier. In addition, we implemented several traditional and deep learning models to serve as baselines for the CNN. Details are as follows.

Classic Baselines

We employed well-known traditional machine learning methods as baselines, including logistic regression (LR), random forest (RF), and support vector machines (SVMs). The implementation used Python language and was based on the scikit-learn toolkit [53].

NN Baselines

We implemented FastText (FAST), RNN, and attention-based bidirectional RNN (ATTENTION) in their original architectures for comparison [56-58]. FAST uses a shallow NN and word representations constructed from bag-of-n-grams, which are averaged into the text representation and input to a linear classifier. RNN considers historical information and shares weights across time by allowing previous outputs to be reused as inputs. ATTENTION builds on RNN by introducing the ability to *attend* to specific subinputs rather than all available information, thereby improving the decisions made at each step.

CNN Model

We implemented a CNN based on Kim Yoon's CNN architecture and used the Keras NN library in Python [59]. The major difference between this implementation and Kim's is that random initialization was used instead of word2vec because both perform equally, max pooling was used instead of global pooling, and filter sizes of 3 and 8 were used instead of 3, 4, and 5. A filter size of 8 was shown to work better with longer documents and would slide over a window size of 8 words at a time [60]. The activation layer following every convolutional layer was the rectified linear unit, $f(x)=max(0,x)$, which maps nonnegatives between the input and feature map. In terms of integrating additional information, domain knowledge features and character2vec (char2vec), into the original CNN system, we simply used concatenation. As the char2vec feature matrix has the same size as the word embedding features and similar logical meaning, we concatenated them together before the main convolution part. Domain knowledge features, a large vector showing the information extracted from the linguistics perspective, were concatenated before the final fully connected layers part. We used a batch size of 512, epoch count of 6, and embedding layer that was learned at the training time. The

embedding layer had a dropout of 0.5 to prevent overfitting and a dimension size of 100. Loss was calculated using cross-entropy. Figure 1 shows our CNN architecture.

Figure 1. Overview of the convolutional neural network architecture.
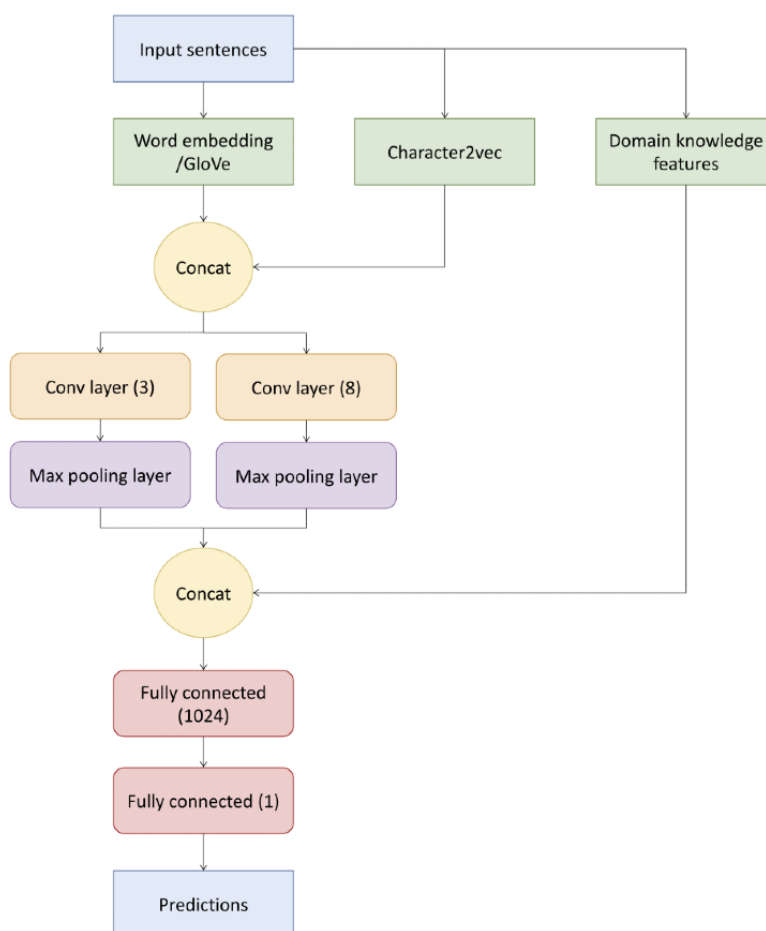
Classification and Prediction

Our hypothesis is from an intuitive standpoint. We considered 2 cases: (1) there exist posts with suicide sentiments among the opiates data set and (2) there exist posts indicative of opioid addiction among the suicidewatch data set. We sought to extract posts that contain both suicidal sentiments and opioid abuse. These 2 cases are referred to as *C1* and *C2* and are represented by their respective models.

Case 1 (C1) is a model that will be trained on data from r/suicidewatch and the control subreddits for positive language, and it will learn to distinguish between suicidal and positive language. As r/opiates contain mixed sentiments, if we feed the out-of-sample data from r/opiates into C1, then C1 should be able to classify between suicide risk posts and nonsuicide risk posts. Posts from r/opiates that are classified as *suicidal* will then contain both suicidality and OUD aspects. In addition, we coarsely assessed the quality of posts predicted as *suicidal* by constructing a word2vec from predictions.

Case 2 (C2) is a model that is trained on data from r/depression and r/opiates. The subreddit r/depression is the most similar to r/suicidewatch which is available. C2 will learn to distinguish

between posts containing drug addiction and no drug addiction. If we feed out-of-sample data from r/suicidewatch into C2, then C2 should be able to classify between posts containing opioid usage and nonopioid usage. Posts from r/suicidewatch that are classified as having opioid usage will then contain both suicidality and OUD aspects.

For classification, the data sets were split with an 8:2 ratio for training and testing. The data for training were split again with an 8:2 ratio for train versus train validation. To evaluate the prediction ability, we sampled 500 posts ad hoc from r/opiates and 500 posts from r/suicidewatch to be annotated. To enforce a lower bound on sample quality and reduce variance among the posts, we only sampled from posts between 30 and 500 words. As class balance is a concern within a sample of 500, a maximum of 250 of 500 posts were selected if they contained a keyword or a key phrase of interest. Thus, the 8:2 train/test split over the entire data set was used for in-sample testing, and then 500 samples were extracted from the 20% of posts that belong to the test split for out-of-sample testing. Table 2 summarizes the C1 and C2 procedures and is illustrated in Multimedia Appendix 2, which demonstrates the workflow for C1.

Table 2. Summary of the 2 models for the 2 cases.

Case	C1	C2
Training and classification		
Goal	Distinguish between suicidal and non-suicidal language; predict for suicidality among opioid users	Distinguish between language of opioid usage and depressed but nonopioid using; predict for opioid usage among suicidal individuals
Data set	51,366 posts from r/suicidewatch and control subreddits	59,940 posts from opioid relevant subreddits and r/depression
Vocabulary size	70,082	64,078
Training	Trained and validated on data from r/suicidewatch and control subreddits	Trained and validated on data from opioid relevant subreddits and r/depression
Prediction on out-of-sample data		
Predicts on	Data from r/opiates	Data from r/suicidewatch
Prediction goal	Predicts for posts containing suicide risk in r/opiates	Predicts for posts containing opioid abuse in r/suicidewatch
Total posts prediction data that are between 30 and 500 words	23,740 posts from r/opiates	21,719 posts from r/suicidewatch
Sample max 250 from data for prediction containing these keywords for MTURK[a]	Commit suicide, suicide, suicidality, suicidal, want to die, want to do, want to overdose	Benzodiazepines, benzos, cocaine, codeine, fentanyl, heroin, hydrocodine, hydrocodone, hydromorphone, hydros, kratom, methadone, morphine, narcotic, narcotics, opiates, opioid, oxycodone, oxycontin, oxycottin, oxycotton, oxymorphone, suboxone
Sample count containing keywords	234	231

[a]MTURK: Amazon Mechanical Turk.

Annotations With Amazon Mechanical Turk

The sampled posts were annotated by crowdsourced workers from MTURK [61]. Each post was annotated by 3 workers, and the final label for a post was determined by the majority rule.

For C1, the MTURK task was titled "Suicidal sentiment detection." The description was "May the user be at risk of suicide or intentional overdose?" with 2 options: "Yes, risk of suicide" and "No risk of suicide." Task keywords included *depression, drug abuse, drugs, mental, mental health, opiates, overdose, suicide, suicide ideation, suicidewatch,* and text. Eligible MTURK workers were required to have masters qualifications and were awarded US $0.04 per annotation. We required no other qualifications. Assuming that the average individual lacks experience in both mental health and drug misuse [62], we defined common slang terms from r/opiates (eg, PAWS stands for postacute withdrawal syndrome) and provided simple instructions based on *Clues to Suicide* by Edwin Shneidman, who is the founder of the American Association of Suicidology and has laid groundwork in the field [63].

For C2, the MTURK task was titled "Opioid addiction among suicidal individuals." The description was "Does the post imply opioid addiction?" with 2 options: "Yes, implies opioid addiction" and "No opioid addiction." Task keywords and qualifications were the same as in the C1 MTURK tasks. We provided a list of opioid names in the instructions.

Results

Comparison of Models and the Impact of Feature Matrices

Metrics were calculated for accuracy, precision, recall, F_1 score, and area under the curve. The complete results can be found in Multimedia Appendix 1. Table 3 provides a subset of the results of comparing the CNN with the traditional baselines using the F_1 score in training the classifiers. Table 4 provides the results of comparing the NNs. At the end of this phrase, the models had been trained and were ready to be used for prediction of out-of-sample data.

Table 3. F_1 scores achieved for the given different combinations of input for classification.

Model	LR[a]	RF[b]	SVM[c]	CNN[d]
C1: r/suicidewatch versus positive control group				
TF-IDF[e]	0.902	0.904	0.915	0.685
word2vec	0.928	*0.927* [f]	*0.943*	0.961
TF-IDF + word2vec	*0.940*	0.921	0.941	*0.963*
TF-IDF + GloVe	0.927	0.829	0.886	0.923
TF-IDF + word2vec + char2vec	0.914	0.790	0.856	0.962
C2: r/depression versus r/opiates				
TF-IDF	0.889	0.800	0.811	0.729
word2vec	0.852	*0.846*	*0.881*	0.961
TF-IDF + word2vec	*0.894*	0.815	0.880	0.965
TF-IDF + GloVe	0.860	0.494	0.765	0.814
TF-IDF + word2vec + char2vec	0.858	0.581	0.741	*0.966*

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]CNN: convolutional neural network.

[e]TF-IDF: term frequency-inverse document frequency.

[f]The best results achieved by the model are in italics.

Table 4. Comparison of text classification neural network baselines with word embedding and convolutional neural network.

Model	FAST[a]	RNN[b]	ATTENTION[c]	CNN[d]
C1: r/suicidewatch versus positive control group				
Accuracy	0.950	0.944	0.939	*0.954* [e]
Precision	0.958	0.953	0.934	*0.968*
Recall	0.957	0.951	*0.965*	0.953
F_1 score	0.957	0.952	0.949	*0.961*
C2: r/depression versus r/opiates				
Accuracy	0.971	0.957	0.969	0.971
Precision	0.964	0.964	0.967	0.970
Recall	0.958	0.923	0.951	*0.962*
F_1 score	0.961	0.943	0.959	*0.966*

[a]FAST: FastText.

[b]RNN: recurrent neural network.

[c]ATTENTION: attention-based bidirectional recurrent neural network.

[d]CNN: convolutional neural network.

[e]The best score for each model is italicized.

From Table 3, we can observe that word2vec contributed to the CNN performance. Another interesting observation is that the classification performance of all models was affected when a pretrained GloVe was used as input. This is possibly due to the local information introduced by using specific subreddits as categories. The introduction of GloVe and character embeddings impeded the performance of RF, which implies that RF experiences more difficulty in learning implicit semantic features than LR and SVM. In contrast, NNs performed well with semantic-based inputs.

Figure 2 provides a visualization of the word embeddings that the NNs trained from scratch. Color intensity and shade from light purple to hot pink indicate increasing emphasis placed on the respective words. CNN, RNN, and ATTENTION share many important words, whereas FastText appears to focus on

different words given its bag-of-N-grams approach but nonetheless achieved high classification performance.

Figure 2. Visualization of word importance as determined from the resultant word embeddings learned from scratch by the neural network models for classification of suicidal vs non-suicidal text. CNN: convolutional neural network; LSTN: long short-term memory; RNN: recurrent neural network.

CNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Fast Text

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

RNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Attention bidirectional LSTM-RNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Evaluating Predictions Using MTURK Annotations

For prediction, we used traditional baselines with word2vec input and NNs. To assess the prediction performance, MTURK was used to obtain labels with 3 annotations per post. C1 predicted for suicidality among opioid data. Annotations were performed by 25 unique workers, with an average of 4 minutes spent on each post. The pairwise agreement among the 3 annotations was 0.632, 0.628, and 0.644. In total, 120 out of 500 posts were ruled as yes, suicide risk. C2 predicted opioid addiction among suicidal posts. Annotations were performed by 92 unique workers, with an average of 11.3 minutes (C1: SD 6.667 minutes; C2: SD 14.633 minutes) spent on each post. The pairwise agreement among annotations was 0.524, 0.604,

and 0.544. In total, 261 out of 500 posts were ruled as *yes, implies opioid addiction*. Table 5 shows the model performance using the majority rule MTURK annotations as the ground truth. Table 5 also shows the differences in performance depending

on the ratio of labels in the data set. To construct a set of data with 500 rows with a certain percentage of suicide risk posts, we sampled using sckit-learn with replace as true.

Table 5. Count and accuracy for model predictions using Amazon Mechanical Turk labels.

Model	LR[a]	RF[b]	SVM[c]	FAST[d]	RNN[e]	ATTENTION[f]	CNN[g]
Count and accuracy for C1 (predicts for suicide risk among opiates data)							
Predicted number of suicide risk	24	12	11	97	93	98	103
All data	0.768	0.750	0.744	0.59	0.608	0.576	0.536
Suicide risk only	0.2	0.1	0.092	0.783	0.791	0.75	0.833
Nonsuicide risk only	0.947	0.959	0.963	0.529	0.55	0.521	0.432
Count and accuracy for C2 (predicts for opioid addiction among suicidewatch data)							
Predicted number of opioid addiction	88	92	105	92	158	110	127
All data	0.524	0.518	0.538	0.54	0.588	0.552	0.562
Opioid addiction	0.230	0.237	0.273	0.251	0.414	0.295	0.334
Nonopioid addiction	0.892	0.869	0.869	0.9	0.806	0.874	0.847

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Observations Regarding C1: Detecting Suicide Ideation in the Context of Opioid Addiction

For both C1 and C2, we noticed that the model metrics changed drastically when the ratio of posts per label was altered. For C1, traditional baselines performed better when there was a mixture of suicidal and nonsuicidal-labeled posts because NNs produce a high number of false positives. However, when only examining suicidal posts, NNs achieved better accuracy. Thus, depending on the domain and data set, it may be better to use models that are superior at the extremes or remain within a relatively less variable range. For our case of predicting suicide risk posts among the out-of-sample opiate data, it is better to err on the safe side than to miss someone who is at risk of suicide. CNN would be a good choice if the sample consisted of many suicide risk posts. The predictive power of the baselines would not be desirable when predicting for posts indicative of suicide risk, as false negatives would be costly in the context of real life.

Given that the models tended to predict at the extremes for C1, we speculated that the models would serve well as weak learners

to an ensemble model. We explored a simple ensemble model using scikit-learn's AdaBoost Classifier and LR, SVM, RNN, and CNN as weak learners. We used MTURK-labeled data of size 500 as the ground truth and the weak learner's probabilistic predictions as the input data. The data were split into training and testing sets at an 8:2 ratio. AdaBoost was run 5 times with different data shuffles and random states, and the average accuracy was taken.

In Figure 3, the top graph illustrates the performance of selected models as we increased the percentage of suicide risk labels from opiate data in C1 from 10% to 90%. The line plot shows the performance of AdaBoost with those models as weak learners. AdaBoost achieved respective accuracies of 0.89, 0.84, 0.82, 0.8, 0.8, 0.86, 0.9, 0.91, and 0.94 for the respective intervals. The bottom graph in Figure 3 displays the model contribution at each step. From the graph, combining the models that predict at the extremes results in a more robust classifier. This is an advantageous approach because the ensemble model now has knowledge about the out-of-sample data, which the weak learner models did not have.

Figure 3. The top bar graph shows changes in prediction performance in C1 depending on data category ratios. The overlaying line plot shows accuracy achieved from using the four models as weak learners in AdaBoost. The bottom graph shows weak learner contribution.
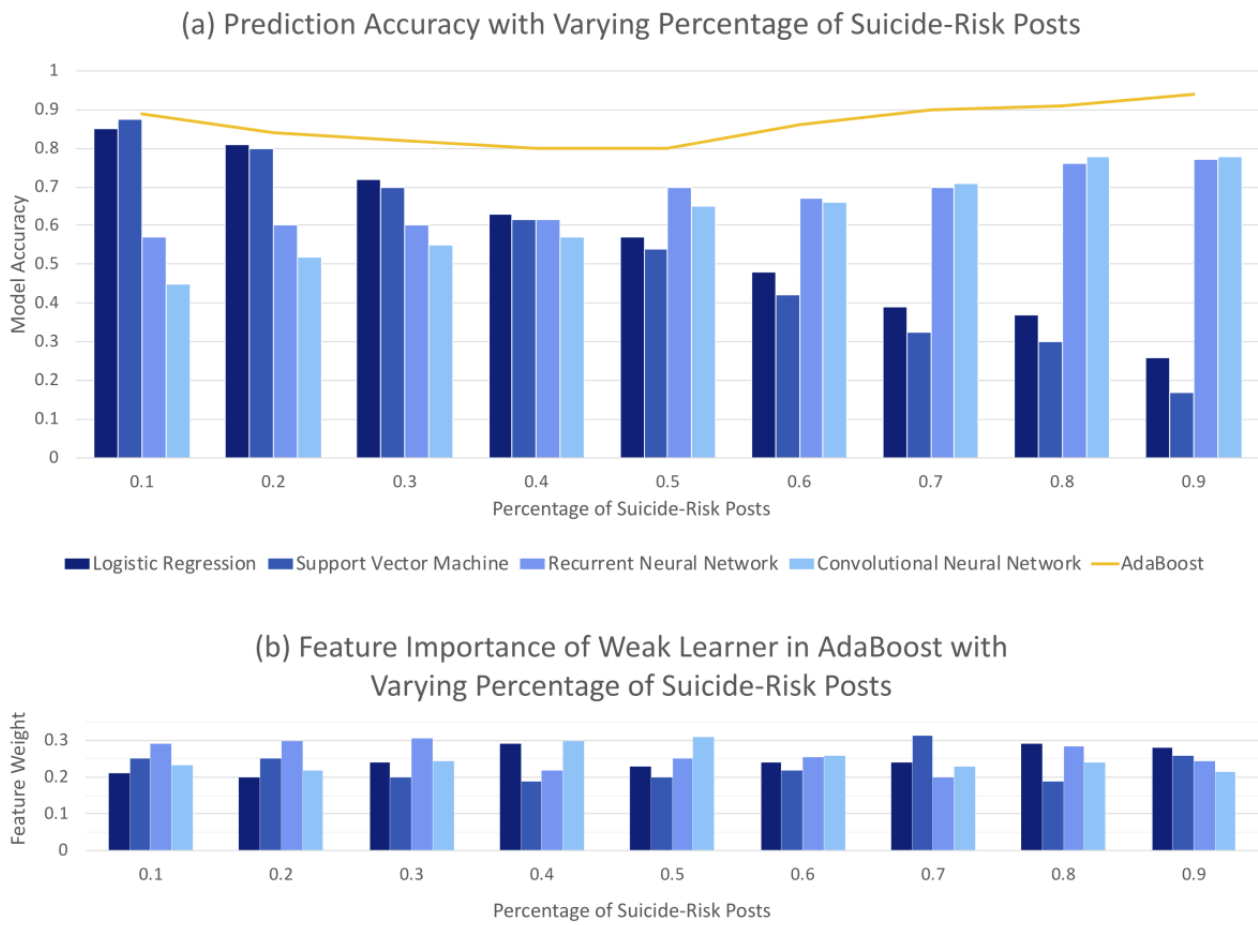
Figure 4 shows the parallel prediction performance for C1 at a more fine-grain level; 50% of each label was sampled for the visual. Each line represents a post. The categorical *x*-axis lists the models, and the intersection of a line with the *y*-axis of a model is the prediction score of the model. The darker lines are MTURK-labeled as indicative of suicide risk (0 on the y-axis), whereas the lighter lines are nonindicative of suicide risk (1 on the y-axis).

Figure 4. Model predictions for individual posts represented as a line. A prediction closer to 1 indicates no predicted suicidal language and a prediction closer to 0 indicates yes predicted suicidal language. Individual posts are color coded by their MTURK label. ATTENTION: attention-based bidirectional recurrent neural network; CNN: convolutional neural network; FAST_TEXT: FastText; LR: logistic regression: RF: random forest; RNN: recurrent neural network; SVM: support vector machine.
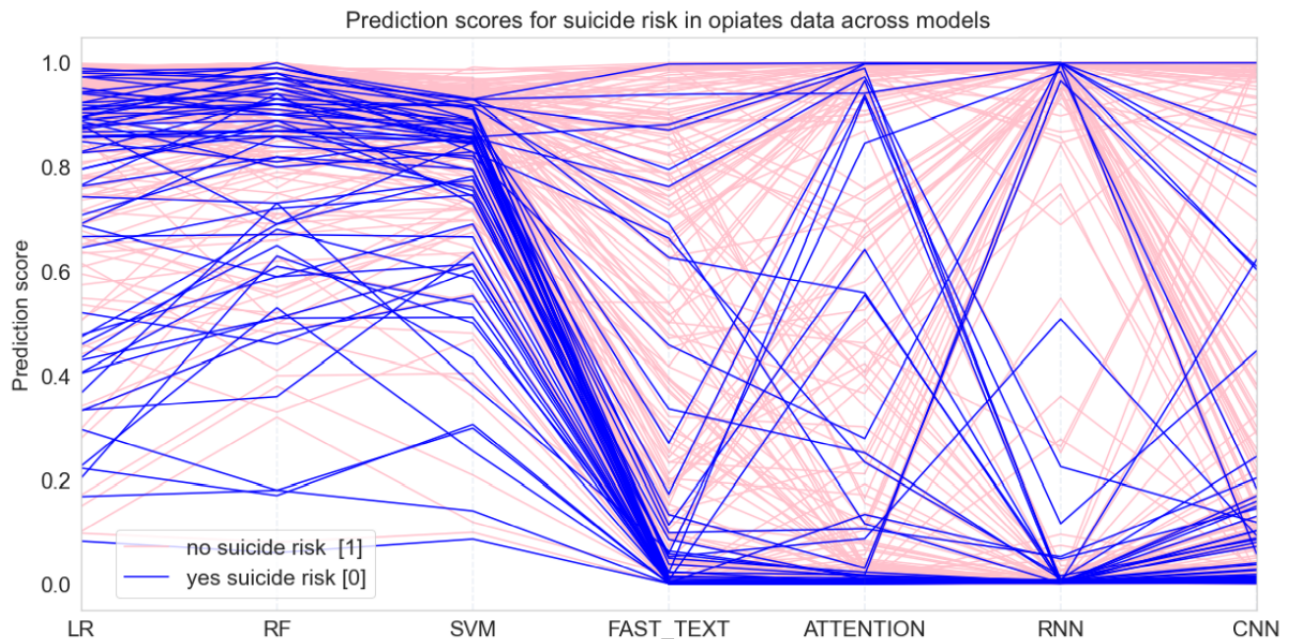
Observations Regarding C2: Detecting Opioid Abuse in the Context of Suicide Ideation

Our hypothesis for C2 was that after being trained on data from r/depression versus the opiates group, the models would be capable of detecting phrases used by individuals who struggle with drug addiction. Although the models all achieved good results in classifying r/depression versus opiate data, they failed to extract opioid-containing posts with respect to the MTURK labels. One possible explanation is that users express themselves differently in different contexts (subreddits). Another possibility is that the language semantics in r/suicidewatch forgo characteristics found in opiates, such as details about withdrawal or relapse that take up the majority of space in any opiate context post. For example, a post in r/suicidewatch may briefly mention a few keywords such as "hydros" or "benzos" in a sentence about overdosing, while the remainder of the post focuses on personal struggles. On the other hand, if a user from r/opiates posts about wanting to commit suicide in r/opiatesrecovery, they will express sentiments regarding the drug because they

know their audience will understand. It is possible that the user is subconsciously assuming that the audience in r/suicidewatch knows little about opioid addiction and therefore omits the consequences of opioid addiction on their life. As a result, an MTURK worker who reads the entire post may observe that the user has easy access to opioids and is experienced with drug mixing; however, the model may fail to pick up such meanings because these exact expressions did not appear in the data it was trained on.

From our new hypothesis regarding language usage in C2, we re-evaluated the C2 predictions. Instead of using MTURK-determined labels, we constructed a set of heuristic labels such that a post is implicative of opioid addiction if the post contains any word from the C2 keywords in Table 2 and is not implicative of opioid addiction otherwise. The findings are shown in Table 6. Although the overall performance was not ideal, all scores increased, supporting our hypothesis that the presence of specific words is important and that MTURK workers' directed attention likely differs from that of the models.

Table 6. Model prediction accuracies with heuristic labels determined by the keyword presence.

Model	LR[a]	RF[b]	SVM[c]	FAST[d]	RNN[e]	ATTENTION[f]	CNN[g]
All data	0.606	0.588	0.648	0.666	0.774	0.69	0.712
Opioid addiction	0.264	0.26	0.347	0.338	0.56	0.403	0.463
Nonopioid addiction	0.9	0.87	0.907	0.948	0.926	0.937	0.926

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Content Overview for Predicted Suicidal Posts in Opiate Data

We constructed a bigram word2vec from posts predicted as *suicidal* in the entire opiate data set (opiate data: n=23,740, as shown in Table 3) to holistically assess the quality of these posts. We then queried for the top 15 most similar words to "suicidal" for each model, as it is more likely to appear with first-person pronouns as opposed to other derivations of the root "suicid." Table 7 summarizes the results from the most similar to least similar words.

Table 7. Top words most similar to suicidal from the subset of opiates data that was predicted to belong to the category suicidal.

Model	Top words
LR[a]	hour, dead, lack, asleep, wake up, sobriety, self, they are, cause, at least, however, hell, later, group, state
RF[b]	easy, living, yet, hit, waiting, probably, cold, the same, such, by, tomorrow, body, constantly, saying, working
SVM[c]	making, everyone, once, pills, without, soon, lol, nothing, around, sorry, thing, withdrawal, start, mental, tolerance
FAST[d]	suicidal thoughts, depressive, extreme, emotional, severe anxiety, existing, depressed, irritable, insomnia, severe, nauseous, mood swings, paranoid, fatigued, having trouble
RNN[e]	severely, depressed, diagnosed with, isolated, unbearable, bipolar, suicidal thoughts, anxious, ptsd, an alcoholic, ocd, overwhelmed, irritable, lethargic, extreme
ATTENTION[f]	diagnosed with, depressive, suicidal thoughts, social anxiety, bipolar, extreme, crippling, irritable, tremors, emotional, borderline, severe depression, brain fog, existing, gad
CNN[g]	paranoid, unhappy, depressed, isolated, apathetic, irritable, an alcoholic, suicidal thoughts, trauma, severely, diagnosed with, brain fog, anxious, manic, emotionless

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Discussion

Overview

The goal of this study is to make use of the structure and data from Reddit to build models that could ultimately predict (C1) suicidal language among opioid users and (C2) opioid usage among individuals with suicidal thinking. We evaluated several models with combinations of semantic inputs and well-known NN text classifiers in their ability to classify and then predict out-of-sample data to extract posts containing both suicide ideation and opioid abuse. MTURK was then employed to provide heuristic ground truths to the out-of-sample data.

Limitations

Owing to the large quantity of data, we assumed that all posts under a subreddit were reflective of the official subreddit purpose (eg, recovering opioid users post in r/opiatesrecovery) and that deviations were neutralized because of the sheer volume of the majority content. A manual review of predicted suicidality among opioid subreddits (C1) found that many cases were wrongly predicted as suicidal because the users were experiencing extreme withdrawal, anger due to relapse, and

even anger due to sobriety because of the inability to use opioids as a coping mechanism. Inclusion of r/offmychest in the control subreddits was an attempt to offset such strong negative emotions that likely do not imply suicidality. A suicidal post may also be predicted as nonsuicidal because of calm resignation language or because the user is writing in self-reflection or to share past experiences.

A manual review of predicted opioid usage in r/suicidewatch (C2) found that poor predictions may have resulted from the sentiment mixture present in the opioid reddits. As the opioid group comprises active users and users in recovery/withdrawal (ie, r/opiates vs r/opiatesrecovery), the language variation could be extreme and vary from anger (ie, *My dealer is ****ing late*) to negative (ie, *I can't find happiness outside of heroin*) to joy (ie, *I've been clean for 70 days!*). All of these posts were given the same label of *opiates* as they were grouped under opioid language. Hence, the presence of drug-specific words is needed to predict opioid usage.

MTURK Limitations

We attempted to overcome MTURK limitations by providing brief context and instructions. Regardless, limitations included lack of domain knowledge, experience with suicide ideation or drug misuse, and subjectivity of interpretation. There is also possible carelessness in annotation, as some posts were relatively long and not all posts may have been read thoroughly. Furthermore, we provided a single post out of context from the user's life; it is difficult to determine how alarming the post is in reality because we cannot know what happened to the user later in time. There was also no way to control how many and which users participated in the assessment—the participation of many distinct workers may have introduced random variations in annotation consistency. Manual review of several posts that were disagreed upon showed that lack of domain expertise can make annotation difficult; for example, a post that was disagreed upon mentioned being *tired of waking up*, in *withdrawal*, and that it is *easier to end it all*, but also mentioned *wanting to get sober*. One MTURK worker may have considered this post as hopeful, whereas another worker may have considered the post to be on the edge of giving up. Despite these limitations, these annotations provide insight into the public perception regarding what is considered suicidal when opiate addiction is involved.

Insights

Studies conducted on opiates and other drug-addicted individuals have shown that opiate addiction is often comorbid with other psychological disturbances, such as personality disorders or trauma [64,65]. We observed this in the simple overview of words most related to *suicidal*. Among the NNs, several personality disorder terms and mental illnesses were extracted, such as *ptsd* (posttraumatic stress disorder), *borderline, bipolar*, and *gad* (general anxiety disorder). Characteristics of withdrawal were also captured: *irritable, insomnia, nauseous*. We can speculate that individuals with personality disorders going into withdrawal phrases are at higher risk of attempting suicide than other subgroups of the population.

Future Directions

There are many possible directions for utilizing Reddit metadata. For subreddits with mixed emotional sentiments such as r/opiates, coarse grain classification can first be applied to discard either too positive or negative posts depending on context. Given the interactivity of Reddit, analyses of comments may show what sort of support is provided by peers who struggle. The peer aspect of Reddit is attractive, and integration of nonjudgmental peer support may help recover OUD users to be more willing to rely on real-life aid. This is one step toward decreasing the frequent social isolation brought upon by OUD.

In general, exploration of Reddit data limits provides many possible study areas, given the amount of active subreddits. Among all posts in r/opiates, we found a handful in which a user who has successfully quit opioid use for many years and returns to relay hope for a full recovery. Such individuals can be seen as *success stories.* It is of particular interest to examine the difference between these individuals and those who end in tragedy, as OUD recovery is difficult and relapse prone [66]. One could also consider the interaction among suicide ideation, OUD, and chronic pain, especially because many OUD cases begin with an individual's or close person's prescription medication [10]. Another focus is the possibility of influencing the OUD individual's behavior; a study in the context of behaviorally targeted advertising for health care based on user searches implies that behavior, as determined by input search terms, is predictive of the individual being influenced by the target advertisement to act, such as visiting a health provider [67]. Therefore, internet behavior may be more closely linked to real-life behavior than expected, and targeted advertising might encourage OUD individuals who are midwithdrawal and fighting relapse to seek help.

The tendency of NNs to achieve high accuracy among purely suicide-labeled posts but perform poorly when the data set is mixed prompts the question of which semantic dimensions are actually being captured. Although we use NNs as a black box in this experiment, future studies may want to focus directly on latent semantic dimensions that may aid in parameter tuning and generalizability. In particular, it would be very interesting to extract the exact features that might generalize across textual data domains by examining the word embeddings that the NNs had trained from scratch to aid in predictions on out-of-sample data [68].

We focused on extracting posts that are indicative of suicide ideation in the context of OUD and did not perform much analysis on extracted posts. Many studies have investigated the content of suicidal individuals, but much fewer studies have been conducted in specific contexts, possibly due to the lack of available data. Gathered predictions are capable of revealing thoughts of suicidal opioid users, despite the limitations of this study. In aiding those recovering from OUD, it may be important to consider beginning at the personal level [69]. Communication between the professional and the patient is important, and personal stories are capable of decreasing stigma by deconstructing stigmatic barriers [70]. As these posts are likely to be raw and honest, they can allow clinical professionals to become more familiar with the mindsets of these individuals.

A less stigmatic perception of patients can lead to patient empowerment, which is necessary in the long term [66,71]. Consider a post that was agreed upon as showing the risk of suicide, summarized quote:

> *as soon as something goes missing, or something goes wrong, you're the first fucking person everyone suspects is at fault...it MUST have been the fucking drug addict, right? ... I may be a dopehead, but I have never been a thief.*

It is easy to advocate understanding and help for those struggling with opioid addiction; however, it is harder to recognize how their own actions may affect those with opioid addiction. We believe that concrete case examples are essential as a first step and that suicide risk assessment should be done on an individual level. Our study seeks to aid in suicide prevention by helping those with OUD gain understanding—our goal is not to substitute for risk assessment using prediction models [72]. Application of gathered insights from the extracted posts can help decrease stigma and clarify wrong assumptions made about drug misusage and suicidal ideation.

Conclusions

The goal of this paper is to make use of big data from Reddit to detect suicidality among opioid users. The structure of Reddit offers categories for data sets, and the social media setting can provide case specifics that may not be obtainable elsewhere. This study can serve as a proof of concept for the use of social media site attributes to aid in machine learning methods and for research directions on the feasibility of NNs to abstract textual data and perform in independent areas. A comparison with a high-performing baseline model implies that the absence of hard-coded features may allow more flexibility and accuracy for models when running on out-of-sample data.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Classification metrics using combinations of different model inputs.
[XLSX File (Microsoft Excel File), 131 KB - jmir_v22i11e15293_app1.xlsx]

Multimedia Appendix 2
Illustration of workflow for C1.
[PNG File , 95 KB - jmir_v22i11e15293_app2.png]

References

1. What is the US Opioid Epidemic? The US Department of Health and Human Services (HHS). URL: https://www.hhs.gov/opioids/about-the-epidemic/index.html [accessed 2020-11-10]

2. Oquendo MA, Volkow ND. Suicide: a silent contributor to opioid-overdose deaths. N Engl J Med 2018 Apr 26;378(17):1567-1569. [doi: 10.1056/NEJMp1801417] [Medline: 29694805]

3. Hedegaard H, Curtin S, Warner M. Suicide Mortality in the United States, 1999-2017. NCHS Data Brief. 2018. URL: https://www.cdc.gov/nchs/data/databriefs/db330-h.pdf [accessed 2020-11-10]

4. Suicide Statistics. American Foundation for Suicide Prevention (AFSP). 2018. URL: https://afsp.org/suicide-statistics/ [accessed 2020-11-10]

5. Norton A. Many Opioid Overdoses May Be Suicides. WebMD. 2018 Apr 25. URL: https://www.webmd.com/mental-health/addiction/news/20180425/many-opioid-overdoses-may-be-suicides [accessed 2020-11-10]

6. Madadi P, Persaud N. Suicide by means of opioid overdose in patients with chronic pain. Curr Pain Headache Rep 2014 Nov;18(11):460. [doi: 10.1007/s11916-014-0460-1] [Medline: 25249422]

7. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Arlington, VA: American Psychiatric Association; 2013.

8. Effective Treatments for Opioid Addiction. National Institute on Drug Abuse (NIDA). 2016. URL: https://www.drugabuse.gov/publications/effective-treatments-opioid-addiction/effective-treatments-opioid-addiction [accessed 2020-11-10]

9. van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev 2010 Apr;117(2):575-600 [FREE Full text] [doi: 10.1037/a0018697] [Medline: 20438238]

10. Han B, Compton WM, Blanco C, Crane E, Lee J, Jones CM. Prescription opioid use, misuse, and use disorders in US adults: 2015 national survey on drug use and health. Ann Intern Med 2017 Sep 05;167(5):293-301. [doi: 10.7326/M17-0865] [Medline: 28761945]

11. Madadi P, Hildebrandt D, Lauwers AE, Koren G. Characteristics of opioid-users whose death was related to opioid-toxicity: a population-based study in Ontario, Canada. PLoS One 2013;8(4):e60600 [FREE Full text] [doi: 10.1371/journal.pone.0060600] [Medline: 23577131]

12. Braden JB, Sullivan MD, Ray GT, Saunders K, Merrill J, Silverberg MJ, et al. Trends in long-term opioid therapy for noncancer pain among persons with a history of depression. Gen Hosp Psychiatry 2009;31(6):564-570 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.07.003] [Medline: 19892215]

13. Hooley JM, Franklin JC, Nock MK. Chronic pain and suicide: understanding the association. Curr Pain Headache Rep 2014;18(8):435. [doi: 10.1007/s11916-014-0435-2] [Medline: 24916035]

14. Hendy HM, Black P, Can SH, Fleischut A, Aksen D. Opioid abuse as maladaptive coping to life stressors in US adults. J Drug Issues 2018 Jul 5;48(4):560-571. [doi: 10.1177/0022042618783454]

15. Sinha R. Chronic stress, drug use, and vulnerability to addiction. Ann N Y Acad Sci 2008 Oct;1141:105-130 [FREE Full text] [doi: 10.1196/annals.1441.030] [Medline: 18991954]

16. Santos DS, Andrade DT, Wan H, Ronzani T, Corrigan P. The Stigmatization of Drug Use as Mechanism of Legitimation of Exclusion. Switzerland: Springer; 2018:15-25.

17. Townsend L, Gearing RE, Polyanskaya O. Influence of health beliefs and stigma on choosing internet support groups over formal mental health services. Psychiatr Serv 2012 Apr;63(4):370-376. [doi: 10.1176/appi.ps.201100196] [Medline: 22337006]

18. Mackey TK, Kalyanam J, Katsuki T, Lanckriet G. Twitter-based detection of illegal online sale of prescription opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]

19. Kalyanam J, Katsuki T, Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]

20. Chan B, Lopez A, Sarkar U. The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. PLoS One 2015;10(8):e0135072 [FREE Full text] [doi: 10.1371/journal.pone.0135072] [Medline: 26252774]

21. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]

22. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. Crisis 2017 Sep;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]

23. Reddit by the Numbers. URL: https://www.redditinc.com/press [accessed 2020-11-10]

24. Johnson GJ, Ambrose PJ. Neo-tribes. Commun ACM 2006 Jan 1;49(1):107-113. [doi: 10.1145/1107458.1107463]

25. Choudhury MD, De S. Mental Health Discourse on Reddit: Self-disclosure, Social Support, and Anonymity. In: The 8th International Conference on Weblogs and Social Media. 2014 Presented at: WSM'14; August 5-8, 2014; Palo Alto, California URL: https://www.semanticscholar.org/paper/Mental-Health-Discourse-on-reddit%3A-Self-Disclosure%2C-Choudhury-De/e58943b8d656fbb07a14f36b28ae211da74992a9

26. DeAndrea DC, Anthony JC. Online peer support for mental health problems in the United States: 2004-2010. Psychol Med 2013 Nov;43(11):2277-2288 [FREE Full text] [doi: 10.1017/S0033291713000172] [Medline: 23410539]

27. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar Lee K, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. Drug Alcohol Depend 2017 Dec 1;181:5-10 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]

28. Ribeiro JD, Joiner TE. The interpersonal-psychological theory of suicidal behavior: current status and future directions. J Clin Psychol 2009 Dec;65(12):1291-1299. [doi: 10.1002/jclp.20621] [Medline: 19827114]

29. Park A, Conway M. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]

30. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018 Sep 9;2018:1-10. [doi: 10.1155/2018/6157249]

31. Sharma E, Choudhury M. Mental Health Supportits Relationship to Linguistic Accommodation in Online Communities. In: The CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 14-16, 2018; Montreal QC, Canada. [doi: 10.1145/3173574.3174215]

32. Kumar M, Dredze M, Coppersmith G, De CM. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. In: The ACM Conference on Hypertext and Social Media. 2015 Presented at: ACM'15; September 20-22, 2015; Guzelyurt, Northern Cyprus URL: http://europepmc.org/abstract/MED/28713876 [doi: 10.1145/2700171.2791026]

33. Grant RN, Kucher D, León AM, Gemmell JF, Raicu DS, Fodeh SJ. Automatic extraction of informal topics from online suicidal ideation. BMC Bioinformatics 2018 Jun 13;19(Suppl 8):211 [FREE Full text] [doi: 10.1186/s12859-018-2197-z] [Medline: 29897319]

34. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz JH, et al. Social media based analysis of opioid epidemic using reddit. AMIA Annu Symp Proc 2018;2018:867-876 [FREE Full text] [Medline: 30815129]

35. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]

36. Handelman LD, Lester D. The content of suicide notes from attempters and completers. Crisis 2007;28(2):102-104. [doi: 10.1027/0227-5910.28.2.102] [Medline: 17722692]

37. Litvinova TA, Seredin PV, Litvinova OA, Romanchenko OV. Identification of suicidal tendencies of individuals based on the quantitative analysis of their internet texts. CyS 2017 Jun 30;21(2):243-252. [doi: 10.13053/cys-21-2-2721]

38. Zhang X, LeCun Y. Text understanding from scratch. arXiv Preprint posted online April 4, 2016 [FREE Full text]

39. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res 2011;12:2493-2537 [FREE Full text]

40. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. 'When 'bad' is 'good'': identifying personal communication and sentiment in drug-related Tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]

41. Husseini OA, Buddhitha P, Husseini OM, Inkpen D. Deep Learning for Depression Detection of Twitter Users. In: The Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLCP'18; June 5-7, 2018; New Orleans, LA p. 88-97. [doi: 10.18653/v1/w18-0609]

42. Kim Y. Convolutional Neural Networks for Sentence Classification. In: The 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: NLP'14; October 23-26, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1181]

43. Husseini OA, Buddhitha P, Husseini OM, Inkpen D. Deep Learning for Depression Detection of Twitter Users. In: The Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLCP'18; June 2-7, 2018; New Orleans, LA p. 88-97. [doi: 10.18653/v1/w18-0609]

44. Johnson R, Zhang T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. In: The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015 Presented at: HLT'15; August 5-8, 2015; Denver, Colorado p. 103-112. [doi: 10.3115/v1/n15-1011]

45. Singh D, Wang A. Detecting Depression Through Tweets. Stanford University. 2016. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6879557.pdf [accessed 2020-11-10]

46. Suicidewatch: Peer Support for Anyone Struggling With Suicidal Thoughts. Reddit. URL: https://www.reddit.com/r/SuicideWatch/ [accessed 2020-11-10]

47. Baumgartner J. Pushshift Reddit API. GitHub. 2017. URL: https://github.com/pushshift/api [accessed 2020-11-10]

48. Mellor T, Boe B. PRAW: The Python Reddit API Wrapper. GitHub. URL: https://github.com/praw-dev [accessed 2020-11-10]

49. Aladağ AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO. Detecting Suicidal Ideation on Forums: Proof-of-Concept Study. J Med Internet Res 2018 Jun 21;20(6):e215 [FREE Full text] [doi: 10.2196/jmir.9840] [Medline: 29929945]

50. Shen J, Rudzicz F. Detecting Anxiety through Reddit. In: The Fourth Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality. 2017 Presented at: CLCP'17; August 3-8, 2017; Vancouver, BC. [doi: 10.18653/v1/w17-3107]

51. Shen J, Rudzicz F. Detecting Anxiety through Reddit. In: The Fourth Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality. 2017 Presented at: CLCP'17; August 5-8, 2017; Vancouver, BC. [doi: 10.18653/v1/w17-3107]

52. Nerbonne J. The secret life of pronouns. What our words say about us. Literary Linguistic Comput 2013 Feb 6;29(1):139-142. [doi: 10.1093/llc/fqt006]

53. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830. [doi: 10.1016/j.patcog.2011.04.006]

54. Rehurek R, Sojka P. Software Framework for Topic Modelling With Large Corpora. In: The LREC Workshop on New Challenges for NLP Frameworks. 2010 Presented at: NLP'10; April 2-9, 2010; Valletta, Malta.

55. Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation. In: The Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]

56. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification. In: The 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017 Presented at: ACL'17; April 3-7, 2017; Valencia, Spain. [doi: 10.18653/v1/e17-2068]

57. Liu P, Qiu X, Huang X. Recurrent Neural Network for Text Classification With Multi-task Learning. In: The Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016 Presented at: CAI'16; April 7-9, 2016; New York, USA. [doi: 10.5555/3060832.3061023]

58. Raffel C, Ellis D. Feed-forward networks with attention can solve some long-term memory problems. arXiv Preprint posted online December 29, 2015. [FREE Full text]

59. Rakhlin A. Convolutional Neural Networks for Sentence Classification. GitHub. 2016. URL: https://github.com/alexander-rakhlin/CNN-for-Sentence-Classification-in-Keras [accessed 2020-11-10]

60.   Zhang Y, Wallace B. A Sensitivity Analysis of (and Practitioners Guide to) Convolutional Neural Networks for Sentence Classification. In: The Eighth International Joint Conference on Natural Language Processing. 2017 Presented at: NLP'17; November 4-9, 2017; Taipei, Taiwan URL: https://arxiv.org/abs/1510.03820

61.   Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2020-11-10]

62.   Tong L, Cheng Q, Homan C, Silenzio V. Learning From Various Labeling Strategies for Suicide-related Messages on Social Media: an Experimental Study. arXiv Preprint posted online January 30, 2017 [FREE Full text]

63.   Shneidman ES. Suicide as psychache. J Nerv Ment Dis 1993 Mar;181(3):145-147. [doi: 10.1097/00005053-199303000-00001] [Medline: 8445372]

64.   DeJong CA, van den Brink W, Harteveld FM, van der Wielen EG. Personality disorders in alcoholics and drug addicts. Compr Psychiatry 1993;34(2):87-94. [doi: 10.1016/0010-440x(93)90052-6] [Medline: 8387417]

65.   Khoury L, Tang YL, Bradley B, Cubells JF, Ressler KJ. Substance use, childhood traumatic experience, and posttraumatic stress disorder in an urban civilian population. Depress Anxiety 2010 Dec;27(12):1077-1086 [FREE Full text] [doi: 10.1002/da.20751] [Medline: 21049532]

66.   Livingston JD, Milne T, Fang ML, Amari E. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. Addiction 2012 Jan;107(1):39-50 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03601.x] [Medline: 21815959]

67.   Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of predicting health care utilization via web search behavior: a data-driven analysis. J Med Internet Res 2016 Sep 21;18(9):e251 [FREE Full text] [doi: 10.2196/jmir.6240] [Medline: 27655225]

68.   Shin J, Madotto A, Fung P. Interpreting Word Embeddings With Eigenvector Analysis. In: NIPS 2018 Interpretability and Robustness for Audio, Speech and Language Workshop. 2018 Presented at: NIPS'18; March 6-10, 2018; Montreal, Canada URL: https://openreview.net/pdf?id=rJfJiR5ooX

69.   Ling W. Prescription opioid addiction and chronic pain: more than a feeling. Drug Alcohol Depend 2017 Apr 1;173(Suppl 1):S73-S74 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.08.643] [Medline: 28363323]

70.   McGinty E, Pescosolido B, Kennedy-Hendricks A, Barry CL. Communication strategies to counter stigma and improve mental illness and substance use disorder policy. Psychiatr Serv 2018 Feb 1;69(2):136-146 [FREE Full text] [doi: 10.1176/appi.ps.201700076] [Medline: 28967320]

71.   Kelly JF, Westerhoff CM. Does it matter how we refer to individuals with substance-related conditions? A randomized study of two commonly used terms. Int J Drug Policy 2010 May;21(3):202-207. [doi: 10.1016/j.drugpo.2009.10.010] [Medline: 20005692]

72.   Large MM. The role of prediction in suicide prevention. Dialogues Clin Neurosci 2018 Sep;20(3):197-205 [FREE Full text] [Medline: 30581289]

Abbreviations

API: Application Programming Interface
ATTENTION: attention-based bidirectional recurrent neural network
CNN: convolutional neural network
FAST: FastText
GloVe: Global Vectors for Word Representations
LR: logistic regression
MTURK: Amazon Mechanical Turk
NN: neural network
OUD: opioid use disorder
PRAW: Python Reddit API Wrapper
RF: random forest
RNN: recurrent neural network
SVM: support vector machine
TF-IDF: term frequency-inverse document frequency

XSL•FO

RenderX

XSL•FO
RenderX

Original Paper

Mapping and Modeling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums: Results of a 15-Year Retrospective Infodemiology Study

Florent Schäfer[1], MSc; Carole Faviez[2], MSc; Paméla Voillot[2], MSc; Pierre Foulquié[2], MSc; Matthieu Najm[2], MSc; Jean-François Jeanne[1], MSc; Guy Fagherazzi[3,4], MSc, PhD; Stéphane Schück[2], MSc, MD; Boris Le Nevé[1], PhD

[1]Innovation Science and Nutrition, Danone Nutricia Research, Palaiseau, France

[2]Kap Code, Paris, France

[3]Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

[4]Center of Research in Epidemiology and Population Health, UMR 1018 Inserm, Institut Gustave Roussy, Paris-Sud Paris-Saclay University, Villejuif, France

Corresponding Author:
Florent Schäfer, MSc
Innovation Science and Nutrition
Danone Nutricia Research
RD 128 Avenue de la Vauve
Palaiseau, 91767
France
Phone: 33 1 69 35 70 00
Email: florent.schafer@danone.com

Abstract

Background: Gastrointestinal (GI) discomfort is prevalent and known to be associated with impaired quality of life. Real-world information on factors of GI discomfort and solutions used by people is, however, limited. Social media, including online forums, have been considered a new source of information to examine the health of populations in real-life settings.

Objective: The aims of this retrospective infodemiology study are to identify discussion topics, characterize users, and identify perceived determinants of GI discomfort in web-based messages posted by users of French social media.

Methods: Messages related to GI discomfort posted between January 2003 and August 2018 were extracted from 14 French-speaking general and specialized publicly available online forums. Extracted messages were cleaned and deidentified. Relevant medical concepts were determined on the basis of the Medical Dictionary for Regulatory Activities and vernacular terms. The identification of discussion topics was carried out by using a correlated topic model on the basis of the latent Dirichlet allocation. A nonsupervised clustering algorithm was applied to cluster forum users according to the reported symptoms of GI discomfort, discussion topics, and activity on online forums. Users' age and gender were determined by linear regression and application of a support vector machine, respectively, to characterize the identified clusters according to demographic parameters. Perceived factors of GI discomfort were classified by a combined method on the basis of syntactic analysis to identify messages with causality terms and a second topic modeling in a relevant segment of phrases.

Results: A total of 198,866 messages associated with GI discomfort were included in the analysis corpus after extraction and cleaning. These messages were posted by 36,989 separate web users, most of them being women younger than 40 years. Everyday life, diet, digestion, abdominal pain, impact on the quality of life, and tips to manage stress were among the most discussed topics. Segmentation of users identified 5 clusters corresponding to chronic and acute GI concerns. Diet topic was associated with each cluster, and stress was strongly associated with abdominal pain. Psychological factors, food, and allergens were perceived as the main causes of GI discomfort by web users.

Conclusions: GI discomfort is actively discussed by web users. This study reveals a complex relationship between food, stress, and GI discomfort. Our approach has shown that identifying web-based discussion topics associated with GI discomfort and its perceived factors is feasible and can serve as a complementary source of real-world evidence for caregivers.

XSL•FO
RenderX

Introduction

Background

Gastrointestinal Discomfort: Prevalence, Impact on Quality of Life, and Management

Gastrointestinal (GI) discomfort (eg, bloating, abdominal pain, constipation) is very common in the general population, with a known impact on well-being [1]. Chronic and severe symptoms of GI discomfort are associated with a significant decrease in quality of life [2]. Irritable bowel syndrome (IBS) is the most studied condition among disorders of gut-brain interactions (DGBIs) [3], with a highly heterogeneous prevalence ranging from 1.1% in France and Iran to 35.5% in Mexico [4,5]. Associated socioeconomic costs are significant because of the important use of health care resources and work absenteeism [6-8]. However, the etiology of DGBIs remains to be poorly understood. Among the pathophysiological mechanisms associated with IBS, GI sensory-motor alterations [9,10], signs of discrete immune dysfunction [11], and increased intestinal permeability [12] are considered important. The possible involvement of gut microbiota in the pathogenesis of GI diseases and the occurrence of GI symptoms has also been explored, as the severity of IBS symptoms is associated with specific intestinal microbiota profiles [13].

Psychological comorbidities are commonly associated with GI symptoms, and the prevalence of anxiety and depression among people with IBS is estimated to be at least two to three times the rate in the general population [14,15]. Concerning women with abdominal pain, cramping, and discomfort, a recent web-based cross-sectional survey study [16] showed that 96 % of women reported that daily activities were disrupted at least sometimes by abdominal pain, cramping, and discomfort and 44 % of women reported that daily activities were disrupted at least often. Other aspects of quality of life, such as quality of work, eating habits, and social activities were also affected in most women [16].

DGBIs are the source of important health care consumption (consultations, complementary examinations and hospitalizations), although difficult to quantify [17], given their chronic nature and the absence of specific diagnostic tests in the case of IBS [18]. In France, a cross-sectional study estimated the average annual direct cost to be €756 (US $888) for one patient and more than 3 days of sick leave per year [17]. Another study conducted in the United Kingdom [19] estimated the total cost of DGBIs in infants to be at least £72.3 (US $93.7) million per year in 2014 to 2015, of which £49.1 (US $63.6) million was the National Health Service expenditure on prescriptions, community care, and hospital treatment.

The role of diet in the pathogenesis of IBS has already been highlighted [20], and food is perceived as a factor of GI discomfort even in the absence of diagnosed allergy or malabsorption [21]. The management of DGBIs especially relies on lifestyle, including physical activity and dietary measures.

Available guidelines [22] recommend regular meal patterns, avoidance of large meals, and reduced intake of fat, alcohol, spicy foods, insoluble fibers, caffeine, and gas-producing foods such as beans, cabbage, and onions. Eating meals in a quiet place (for about at least 20 min, without working) with sufficient chewing and hydration (1.5 to 2 liters per day) is also recommended [22]. Dietary interventions (probiotics, prebiotics, and synbiotics) and restriction diets (eg, low-fermentable oligosaccharides, disaccharides, monosaccharides, and polyols, also known as low-FODMAP diet) have also been explored as potential therapeutic solutions in IBS [23]. Available pharmacological treatments targeting either the GI tract or the brain have also shown some therapeutic value and include antidepressants, prokinetic agents, and painkillers [24].

Social Media as a Real-World Health Data Source

The penetration of social media into modern society has become a global cultural phenomenon. Patients use peer-to-peer virtual communities and social media to share their experiences regarding their treatments and diseases. The use of social media allows large groups of people to create and share information, opinions, and experiences about health conditions and medications through web-based discussion [25]. Social media can therefore be considered as a new data source to assess population health and quality of life, understand adherence to treatments, or identify adverse drug reactions. Patients highlighted the benefit of web-based interactions with other patients. For example, sharing information through social networks enabled patients to better communicate with health care providers. Patients often use social media to discuss drug side effects, quality of life and adherence to therapies. To analyze such data, which can be voluminous, appropriate tools are needed. Text mining techniques allow the classification and summarization of text data such as messages [26]. This set of techniques has been used to extract information from electronic health records [27,28]. They have also been used for various use cases with social media data. Some authors [29] studied the messages from patients with breast cancer treated with aromatase inhibitors. In France, some studies have been published on the misuse and pharmacovigilance signals of methylphenidate [30,31], the incorrect use of neuroleptics regarding anxiety [32], the safety profile of Levothyrox and the dynamics of its reporting on social media during the summer of 2017 [33]. In multiple therapeutic areas, including diabetes and obesity, social media has been considered a real-world health evidence data source [34]. Although web-based discussions are unstructured as compared with conventional clinical data, their volume (hundreds of thousands of users) is very important when compared with clinical sets, and this information can therefore be considered as a complementary source of health data in observational research.

Objectives

This study was designed to explore perceived GI discomfort and better understand its determinants, on the basis of a retrospective assessment of web-based social media posts, which

we considered as a real-life source of health information. In this study, we aimed to answer 3 main research questions: (1) Can we identify topics discussed by web users reporting symptoms of GI discomfort? (2) Can we categorize these users on the basis of the reported symptoms of GI discomfort and level of activity on social media while considering their age and gender? (3) Can we identify the perceived factors causing GI discomfort as reported by web users?

Methods

Data Sources and Data Extraction

Messages were retrieved (Figure 1) from general and specialized French medical web forums. Only messages from publicly available sources were extracted [35]. Messages published between January 2003 and August 2018 containing keywords related to GI discomfort were retrieved along with messages from 3 GI discomfort–related subforums of *Doctissimo* [36]: (1) *constipation, other transit disorders*; (2) *digestion, heartburn, gastroesophageal reflux disease (GERD);* and (3) *abdominal pain, stomachache and ulcers*. Messages were automatically extracted using the published *Detec't webcrawler* [37,38] developed by Kap Code. A web crawler is an engine that browses through hyperlinks and stores them for future download of associated web pages (identified by the visited hyperlinks) [39]. Scraping of messages was performed according to the HTML structure of each forum. All discussions containing at least one of the keywords or one of their synonyms were automatically retrieved with all the associated metadata, deidentified and cleaned (signature and quote withdrawal) before being stored in a study-specific database. A description of the extracted corpus is presented in Multimedia Appendix 1. A complete list of the forums that were crawled is presented in Multimedia Appendix 2, and a list of the keywords used for message retrieval is detailed in Multimedia Appendix 3.

Figure 1. Study framework.
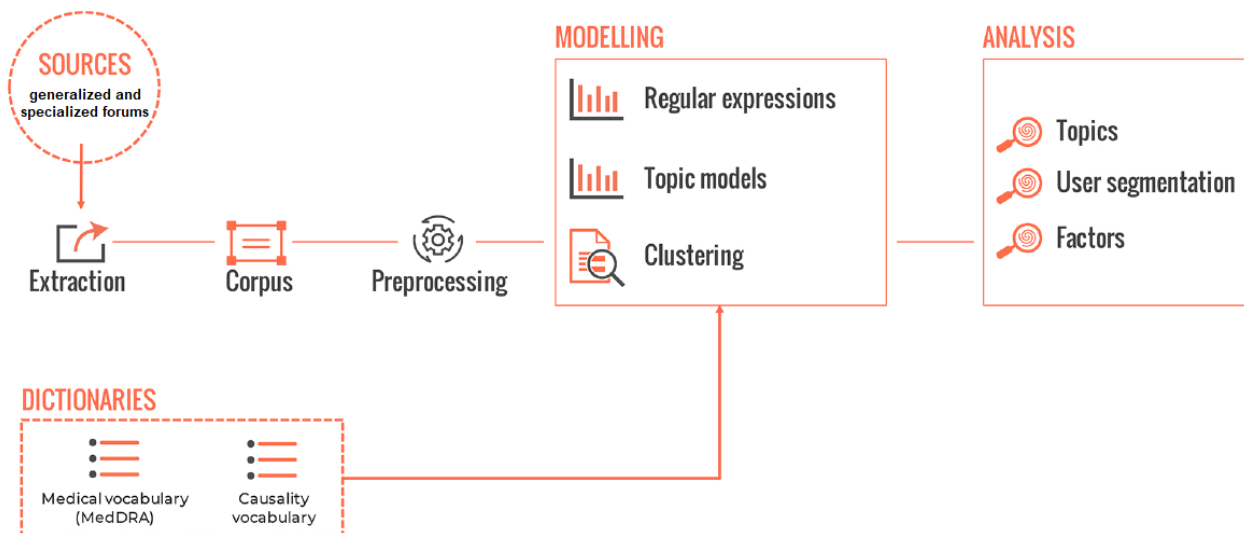
The corpus is presented in number of messages, one message being one statistical unit. As the total number of extracted messages could not be estimated in advance, no a priori assumption was made about the topics being discussed by web users, the clusters to be segmented, or the perceived factors to be identified, and no sample size was calculated for this observational study.

Data Exclusion

The analysis corpus consisted of the corpus cleaned after the removal of messages containing predetermined keywords written in a language other than French, messages containing at least one of the study-specific exclusion words (such as animal-related vocabulary or GI symptoms being used out of context), messages coming from specific URLs and duplicates, as presented in Figure 2.

Figure 2. Flowchart presenting message extraction.
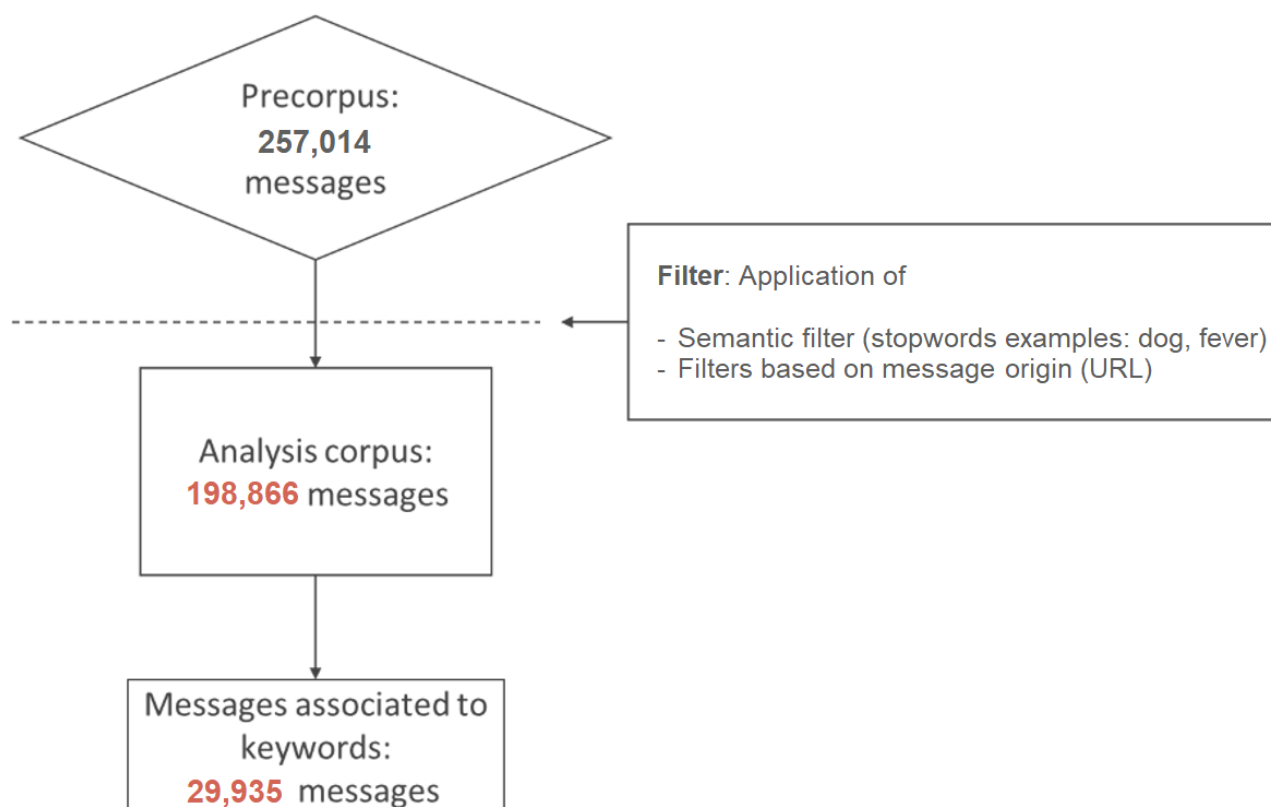
Statistical Analyses

Discussion Themes and Topics

A topic model was applied to identify the themes addressed in the messages. Topic models consist of text mining approaches that aim to automatically identify the abstract themes addressed in a collection of documents. The simplest and most current form of topic models is latent Dirichlet allocation (LDA) [40,41]. It is based on the hypothesis that each document in the corpus corresponds to a distribution of several topics, these distributions being Dirichlet prior. The modeled topics are probability distributions over the tokens (words or a sequence of several adjacent words) found in the corpus. There is no prior assumption made about the nature of topics present in the corpus under study. These models have already been used to analyze health-related topics within tweets [40,41] or online forums [29,42,43].

For this study, the correlated topic model was used [44,45]. In addition to being based on LDA [44], it considers the existing relations between discussed topics as an additional parameter. The estimated correlation between 2 topics indicates the extent to which these 2 topics emerged simultaneously in posts.

The modeling of the studied corpus went through different steps so that the topic model could be applied [30]. The model was estimated using a variational expectation maximization algorithm [44,45], which approximates the posterior distribution of topics on the corpus by finding the best combination of variational parameters. Topics being probability distributions over tokens of the study corpus, they can be characterized by the highest per-topic probability tokens. Weighting these

probabilities through term frequency-inverse document frequency (TF-IDF) allows the allocation of higher importance to topic-specific tokens [45]. In this case, the per-topic probability of a token was weighted by the inverse of the probabilities of this token in other topics. For each topic, tokens were ranked from highest to lowest weighted TF-IDF value of their probability in this topic [45]. For each topic, the first 15 tokens obtained through this ranking were considered the most associated tokens. These were defined as the set of characteristic tokens and used to label each topic. This label should be a synthesis of the characteristic tokens expressed. Correlations between the different topics were measured. Topics were considered associated when correlations were higher than 0.2 in absolute value. This threshold has been set empirically to allow a post to be associated with 5 topics at most. Discussion topics were merged in *groups* of topics on the basis of the values of correlations, and some focus was on categories of interest by applying a new topic model to the associated messages. The analysis was performed using the Structural Topic Model package [46] with R environment version 3.5.2.

User Segmentation

A nonsupervised clustering algorithm (agglomerative hierarchical clustering) was applied on the data to categorize users according to their activity profile, using 36 different features, from 3 categories: symptoms of GI discomfort, identified topics, and website activity. These categories and features are presented in Multimedia Appendix 4.

A specific list of symptoms related to GI discomfort was established on the basis of the Medical Dictionary for Regulatory Activities (MedDRA) terms and colloquial language [37]. For

this, a review of the medical dictionary, MedDRA version 15.0, was performed to identify all the terms that may be associated with GI discomfort. Subsequently, these terms were manually grouped by anatomical region or pathophysiological mechanism (esophageal disorders, gastric disorders, GI disorders, pain, appetite disorder, etc). A list of these regions and mechanisms used to group these messages is presented in Multimedia Appendix 5. Fifteen different categories of symptoms were established. A manual enrichment of these groups was made using colloquial language. Automatic screening of messages allowed the identification of specific GI symptoms expressed by web users.

Website activity of users was measured through different features such as the number of messages, the number of discussions, the dates of the first and last post, the forum name, or the mean posting span. Proportions of posts associated with categories of topics identified for the first objective were considered as the last type of features to describe web users. Created clusters of users were described via identity cards presenting the features that allowed to single them out with their age and gender distribution.

Web users' gender was determined through the identification of regular expressions (gendered past participles, adjectives, and names) in messages and the application of a support vector machine on the basis of message content. This method achieves 88% accuracy and is the subject of a pending publication. Web user age categories were identified on the basis of the use of regular expressions of the author's age in the messages, such as *J'ai 45 ans* (*I am 45 years old*). Each pseudo was associated with one gender (male, female, or unknown) and one age category (20 years or younger, 21-30 years, 31-40 years, 41-50 years, 51-60 years, 61 years or older, and unknown). The generated identification cards (features, age, and gender) were used to characterize each cluster and evaluate whether these characteristics were homogenous between clusters.

Factors of GI Discomfort

Factors perceived as responsible for the reported symptoms of GI discomfort were identified using a mixed automated analysis method combining syntactic analysis and topic model. The syntactic analysis was designed to identify (1) the messages containing extracted keywords and a causality term and (2) the sentences and phrases where the causality terms are present. To identify causality, a specific dictionary made of terms associated with causality was created. Causality terms consisted of terms or groups of terms expressing causality in French. These terms could be verbs conjugated at different times and pronouns (eg, *me donne* [makes me], *lui provoquait* [caused/triggered], *entrainent* [lead to/cause], etc), prepositions (eg, *à cause de* [because of/due to], etc), and conjunctions (eg, *dès lors* [since/consequently], *du fait de* [given that], etc).

Depending on the causality terms, the position of the segment of phrases where factors could be identified was located either before or after the term, as presented in Figure 3. A topic model was then applied to the sections of messages depending on the direction associated with each causality term. These terms are presented in Multimedia Appendix 6 along with the associated direction of the segment of phrases to be analyzed. The number of topics was set to 30 to maximize the number of topics associated with factors that could arise. These topics were reviewed manually, and the topics of interest related to factors were identified.

Figure 3. Example of messages. (A) Causality term associated with right section. (B) Causality term associated with left section. The sections in which the topic model is applied are indicated in green.
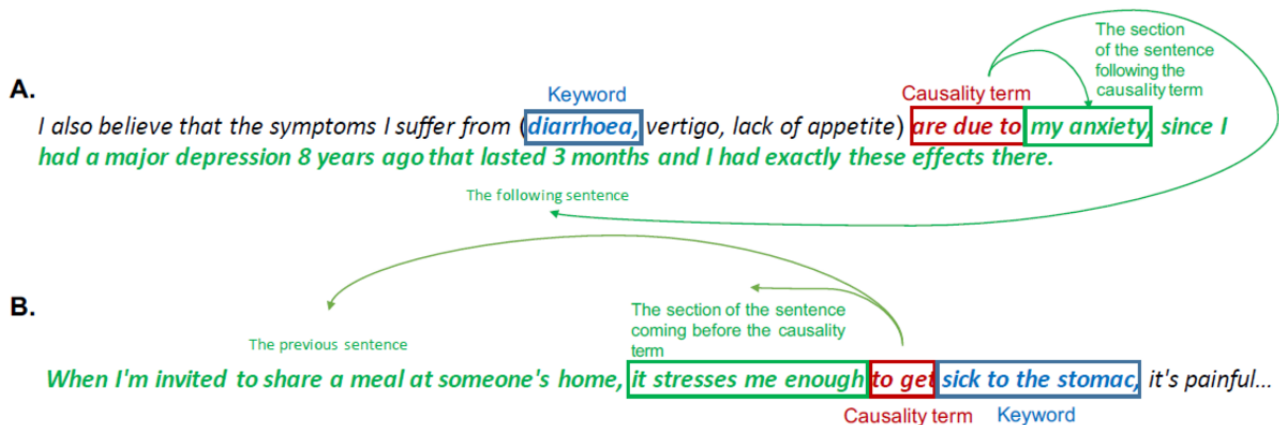
Results

Data Set Description

After cleaning and formatting, the obtained corpus contained 198,866 messages. A total of 36,989 different web users were associated with this corpus. A total of 29,935 messages (corresponding to 16,746 different web users) contained at least one of the extraction keywords, and 181,365 messages came from the *Doctissimo* subforums (Multimedia Appendix 2).

The most frequent keyword was *abdominal pain,* as presented in Table 1. The most frequently mentioned keywords were lay vocabulary: *nausea*, *colic*, *vomiting* and *diarrhea*. More expert terms such as *irritable bowel syndrome*, *irritable bowel,* or *dyspepsia* were used less frequently.

Messages were retrieved from 14 different generalized and specialized web forums. Extracted data mostly came from *Doctissimo* (182,647/198,866, 91.84% of messages; 27,415/36,989, 74.12% of users). The most frequently used data sources were *Aufeminin* (2325/36,989, 6.29% of web users), *Sante-medecine* (1375/36,989, 3.72% of web users), *Atoute.org* (1350/36,989, 3.65% of web users) and *Onmeda* (1341/36,989, 3.63% of web users).

Table 1. Most frequently used extraction keywords.

Keyword extraction (top 20)	English translation	Number of messages, n (%)
Mal au ventre	Abdominal pain	9011 (4.53)
Nausée	Nausea	2571 (1.29)
Colique	Colic	1795 (0.90)
Vomissement	Vomiting	1722 (0.87)
Gargouillis	Borborygmi	1698 (0.85)
Diarrhée	Diarrhea	1642 (0.83)
Ballonnement	Bloating	1548 (0.78)
Constipation	Constipation	1352 (0.68)
Des gaz	Gas	1070 (0.54)
Rot	Burp	996 (0.50)
Pet	Fart	847 (0.43)
Colopathe	IBS[a]	837 (0.42)
Colopathie fonctionnelle	IBS	757 (0.38)
Côlon irritable	Irritable bowel	609 (0.31)
Problèmes intestinaux	Bowel problems	599 (0.30)
Reflux gastrique	Acid reflux	488 (0.25)
Selles molles	Loose stools	452 (0.23)
Chiasse	Runs	402 (0.20)
Flatulence	Flatulence	364 (0.18)
Dyspepsie	Dyspepsia	338 (0.17)

[a]IBS: irritable bowel syndrome.

Discussion Themes and Topics

A total of 18 topics of interest were identified on the basis of manual labeling and review of the data (Table 2). Although the most discussed topic was related to everyday life, the second most discussed topic was related to diet.

Topics were gathered into 6 main groups of clusters on their correlations: *consultations, diet, symptoms, quality of life, treatments* and *stress and symptoms*. The *symptoms* category was further subdivided into 3 subcategories: *abdominal pain, GERD,* and *digestion*. The 8 derived categories were used for user segmentation.

A second topic model was applied to the messages from the group of topics *diet*. This focus allowed the identification of a constellation of subtopics related to symptoms associated with diet (*nausea and vomiting, bloating and gastric reflux*), to the importance of adapting diet to avoid troubles (*diet as a solution to gastric troubles, recipes, balance in diet and efficiency of modifying the diet*), and to food intolerance *and intestinal microbiota (gluten, dairy products and intestinal flora)*. This subtopic was rising in 2017 (ie, the relative number and number of posts discussing this topic), which is the last complete year of the analysis corpus.

Another topic model was applied to messages from the group of topics *stress and symptoms*. The main identified subtopic was addressing solutions to stress (*sport, courage* and *anxiety*). Other identified subtopics revealed a complex relationship between stress and symptoms of GI discomfort, as some subtopics were presenting GI symptoms as a cause of stress (*impact on social life and persistent gastric symptoms*) and stress as a cause of GI symptoms (*because of stress and pain because of problems*), sometimes during specific periods (*GI symptoms flare during exams or depending on the menstrual cycle*).

Table 2. List of modeled and merged topics.

Topics	Number of messages, n (%)	Number of users, n (%)	Group
Prediagnosis medical consultations	4752 (2.39)	2095 (5.66)	Medical consultations
Examinations for diagnostic purposes	3183 (1.60)	1283 (3.47)	Medical consultations
Postdiagnosis medical consultations	2795 (1.41)	1422 (3.84)	Medical consultations
Medical examinations	6852 (3.45)	4013 (10.85)	Medical consultations
Diet	12,802 (6.44)	4727 (12.78)	Diet
Food and IBS[a]	2211 (1.11)	988 (2.67)	Diet
Abdominal pain and nausea	12,385 (6.23)	6939 (18.76)	Symptoms—abdominal pain
Abdominal pain	8370 (4.21)	4130 (11.17)	Symptoms—abdominal pain
Gastroesophageal reflux disease	2001 (1.01)	672 (1.82)	Symptoms—GERD[b]
Gastroesophageal reflux disease and heartburn	6202 (3.12)	2337 (6.32)	Symptoms—GERD
IBS[a]	2538 (1.28)	1061 (2.87)	Symptoms—Digestion
Digestion	12,521 (6.30)	5290 (14.30)	Symptoms—Digestion
Digestive disorders in children	4520 (2.27)	2216 (5.99)	Symptoms—Digestion
Impact on everyday life	8628 (4.34)	3672 (9.93)	Quality of life
Everyday life	16,176 (8.13)	5902 (15.96)	Quality of life
Medication efficiency	6240 (3.14)	2825 (7.64)	Treatments
Information about the treatments	1561 (0.78)	856 (2.31)	Treatments
Stress and symptoms	5679 (2.86)	2971 (8.03)	Stress and symptoms

[a]IBS: irritable bowel syndrome.

[b]GERD: gastroesophageal reflux disease.

User Segmentation

The algorithms based on regular expressions identified the gender for 14,441 users and the age for 4802 users. These results are presented in Table 3. The sex ratio was 0.20.

Table 3. Users' characteristics: number of web users and relative number of web users (among 36,989 users).

Age range (years)	Women, n (%)	Men, n (%)	Unknown, n (%)
0-20	569 (1.54)	117 (0.32)	715 (1.93)
21-30	997 (2.70)	195 (0.53)	691 (1.87)
31-40	466 (1.26)	87 (0.24)	272 (0.74)
41-50	227 (0.61)	41 (0.11)	128 (0.35)
41-60	123 (0.33)	19 (0.05)	69 (0.19)
61 and over	42 (0.11)	13 (0.04)	31 (0.08)
Unknown	9647 (26.08)	1898 (5.13)	20,642 (55.81)

A total of 12.98% (4802/36,989) of the users were characterized according to their age range (62,146/198,866, 31.25% of the messages), and 39.04% (14,441/36,989) of the users were characterized according to their gender (118,882/198,866, 59.78% of the messages). The nonsupervised, bottom-up, hierarchical clustering exhibited 16 different clusters according to the segmentation features presented in Multimedia Appendix 5. After clustering, the expression of these features enabled the visual identification of the expression of the features in a heatmap, which is presented in Figure 4. A review of this heatmap enabled the manual identification of clusters that are presented in Table 4. These clusters were labeled according to the expressed features, and only clusters of more than 100 web users were considered. Clusters are generally characterized by types of symptoms or diagnosed diseases (GERD, digestive disorders, stress and abdominal pain).

Figure 4. Heatmap presenting the results of the hierarchical clustering of web users on the basis of the discussed topics, symptoms, and activity on websites. GERD: gastroesophageal reflux disease.
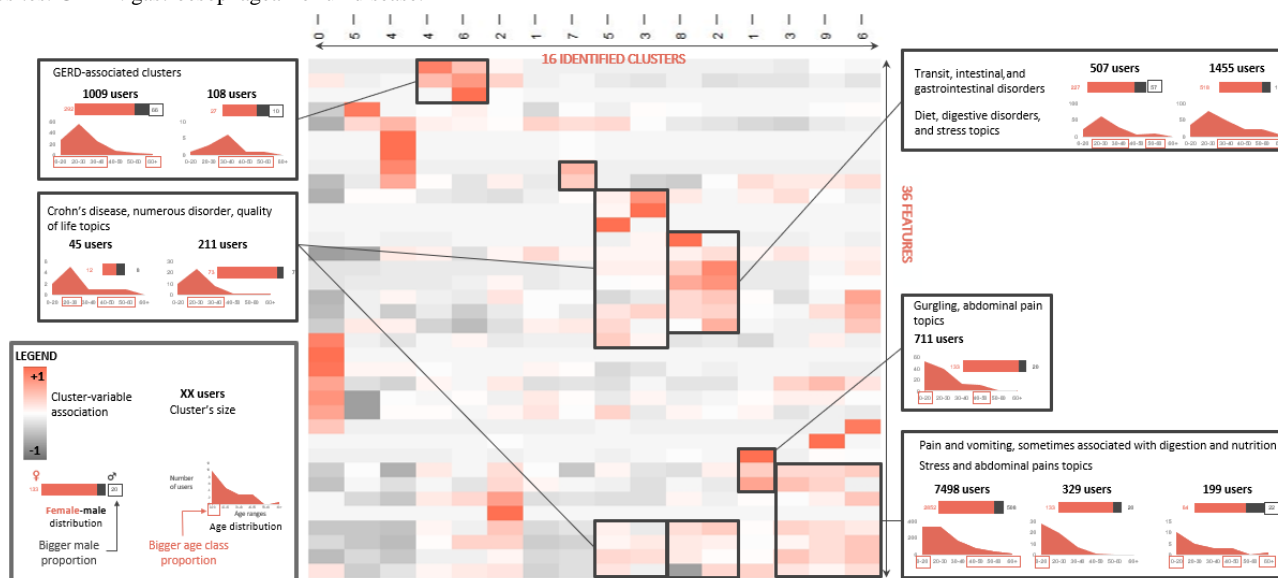
Table 4. List of main clusters of web users that were identified (36,989 users).[a]

Cluster name	Number of users, n (%)
Pains and vomiting, stress and abdominal pain	8026 (21.70)
Gastro intestinal disorders associated with diet, digestive disorders, and stress	1962 (5.30)
GERD[b]	1117 (3.02)
Borborygmi and abdominal pain	711 (1.92)
Crohn disease	256 (0.69)

[a]Clusters were named according to the features that were expressed in users' messages.

[b]GERD: gastroesophageal reflux disease.

Two clusters were strongly associated with GI disorders. These 2 clusters were also associated with the groups of topics, *stress, symptoms,* and *diet*. Three clusters were strongly associated with pain, vomiting, and groups of topics: *stress and symptoms* and *symptoms—abdominal pain*. These clusters were associated with a younger population. More generally, clusters associated with undiagnosed symptoms were often associated with the *Stress* topic, whereas clusters associated with diagnosed diseases (GERD and Crohn disease) seemed to be associated with fewer stress features. The 4 clusters associated with the *Diet* topic were also associated with the group of topics *stress and symptoms*.

Factors of GI Discomfort

The causality dictionary contained 170 terms. These causality terms were searched in the analysis corpus subset containing extraction keywords (29,935 messages). A total of 20,500 messages (corresponding to 10,848 users) were identified, and a new topic model was applied. The characteristic tokens and a sample of characteristic messages associated with each topic were manually reviewed to identify the themes addressed and the topics related to the causes of GI discomfort. In total, 10 topics of interest arose and were manually labeled and grouped according to the type of factors they expressed. This led to the identification of 7 different types of factors. The proportion of messages associated with these factors was calculated (Table 5). The factors that were mostly perceived by web users were related to the *psychological context* (psychological and social factors), followed by *diet* (nutritional factors and allergens or food intolerances) and *medical factors* (GI diseases, gynecological factors, and medical complications).

Table 5. Perceived factors of gastrointestinal discomfort. Proportions are calculated among the messages with causality terms (20,500 messages).

Factors and topics	Messages, n (%)
Psychological factors	4327 (21.11)
Stress	2548 (12.43)
Anxiety	2374 (11.58)
Nutritional factors	3224 (15.73)
Allergens	2857 (13.94)
Diagnosed gastrointestinal diseases	2697 (13.16)
Digestive disorders	1516 (7.40)
Medical examinations	1296 (6.32)
Gynecological factors	1898 (9.26)
Obstetrical factors	1005 (4.90)
Gynecological factors	980 (4.78)
Social factors	1568 (7.65)
Medical complications	1070 (5.22)

Discussion

Principal Findings

GI discomfort is actively discussed in French web forums, as shown in this study, which enabled the identification of 198,866 messages. In a subcorpus of web users who we were able to characterize according to age and gender, mostly women aged below 40 years were represented (Table 3). The gender and age distribution of most active users tended to mirror the higher prevalence of DGBIs such as IBS in younger women [47].

Users described how they were adapting their diet to avoid symptoms linked to perceived food intolerance (*gluten and dairy products*) associated with gut microbiota (*intestinal flora*). This is in line with the increasing number of reports in the literature about the controversial concept of nonceliac gluten sensitivity [48].

As stress was reported by users as both the cause and consequence of their GI symptoms, a focus on the *stress and symptoms* group revealed discussion topics related to the impact of stress on quality of life and solutions to reduce stress, such as physical activity. The role of psychological factors such as anxiety in eliciting or worsening GI symptoms is well established in the literature, both in the general population and in patients affected by DGBIs [49].

User segmentation led to the identification of 16 different classes, grouped into 6 main clusters. The classes associated with symptoms were generally associated with stress, with a stronger association in the case of abdominal pain. The 3 clusters associated with abdominal pain (attributed to diet or digestion) were associated with a younger population. In addition, in all groups of users reporting diet-related features, *Stress* and *Symptoms* topics were also expressed. This suggests a complex relationship between diet, stress, and symptoms of GI discomfort in a real-life setting. We believe that these results may appeal to researchers collecting dietary parameters in nutrition and clinical studies, as the monitoring of dietary intake and habits is important in prospective medical research studies [50]. Indeed, further context on meal intake (such as social and emotional context) should be collected to ensure that eating behavior and associated sentiments are accounted for. In a recent review, it was highlighted that emotion tracking is a lacking feature in most downloaded smartphone apps that are used for dietary assessment [51]. However, these tools could include features that may be used to examine emotions associated with meals in an observational setting at the population level. Such features would also be important to obtain further information on background diet, which is important when evaluating the efficacy of food and dietary interventions in research [52,53], especially in patients with DGBIs [54].

Our analysis identified 7 categories of factors of GI discomfort (psychological, nutritional, allergens, diagnosed GI diseases, gynecological, social, and medical complications), showing that food and psychological factors are perceived by web users as the major causes of GI discomfort. Identification of perceived factors revealed complex associations between food and health parameters. As an example, the use of *fibres* (fibers) keyword revealed a contrasted perception by web users, fibers being seen both as a solution to and cause of GI discomfort (associated with nutritional factors). As noted by another research team analyzing bowel disease–related tweets [55], web-based messages about foods and diet (in this case fiber, iron and magnesium) can be positively or negatively perceived depending on the conditions of web users. Regarding other keywords representing factors of food origin, milk, gluten, and fruits were some of the most frequently used terms.

Comparison With Prior Work

The results of another study aimed at characterizing the inflammatory bowel disease community based on Twitter discussions during an 8-month period were published while this paper was being prepared [55]. This research team also identified that web users shared their experiences and looked for medical advice and that users' discussions were mainly about inflammatory bowel disease symptoms, related diseases

(including anxiety disorders), and foods and diets (including dietary interventions, such as gluten-free and probiotics). These findings are consistent with the main discussion topics that we identified in our corpus, even though the media source (tweets) and language (English) were different, and the studied indication (inflammatory bowel disease) was more specific for this study.

Recently, the smartphone app, *My Symptoms* [56], was completed by 163 participants to track food intake, psychological distress and GI symptoms in a research study aiming to identify the associations between these parameters. The results of this study were recently published by the research team [57], which described strong symptom-symptom associations, especially abdominal pain, bloating, gas-related discomfort, and psychological distress. All these parameters are topics or subtopics that were identified in our study; at the same time, we also noted an association between abdominal pain and stress.

We identified that topics related to medical consultations and medical examinations were frequently discussed, suggesting an important use of the health care system due to GI disorders, which is consistent with prior work [17]. When this paper was drafted, results of another study relying on another source of real-world information, the French National Health Data System (*Système National des Données de Santé*) [58], were published [59]. This study aimed to assess health care use in a specific case of IBS. This study also revealed an important use of the health care system by patients with IBS, also interestingly suggesting an important medical nomadism for these patients in France.

Limitations

A limitation of this study is inherent to the particularities of web forums where web users do not necessarily reflect the characteristics of the general population. Although the important number of extracted messages could favor the variability of users' characteristics, these results cannot be generalized to all patients affected by GI discomfort.

An extraction bias is associated with the considered data sources and keywords selected for analysis. Moreover, information found in messages cannot be interpreted as it would be from a questionnaire. Handling missing information is a key example: the fact that a piece of information is not found in messages does not mean that users did not experience it. For these reasons, it can be difficult to draw conclusions in cases of missing or unclear data. Another example is the identification of the age

and gender of web users, which is not possible if not indicated in the source and not systematically identified in our study.

An additional limitation is the observational bias inherent to semantic analysis and natural language processing. The use of automatic analysis allows us to analyze a large amount of information but is subject to limitations arising from the abilities of the algorithms. Moreover, regarding topic models, the fact that topics must be manually labeled is also a source of bias.

The processing of lay language as source data prevents us from drawing further conclusions on the identified factors of GI discomfort that would require a high level of knowledge from web users. These factors are, therefore, presented as *perceived* factors in this paper as the assessment of their relationship with symptoms of GI discomfort results from self-assessment by web users. In addition, this analysis revealed several misconceptions, especially about factors of food origin. For example, web users may discuss food-allergic reactions but may refer to symptoms that are unlikely to be mediated by the immune system or compounds that are not known to be the cause of allergic reactions but rather of hypersensitivity or intolerance (eg, lactose). This is consistent with the results of a population-based survey published during this study, concluding that population-estimated prevalence of allergy was twice as important as the one estimated by physicians [60]. However, despite these limitations due to the analyses in lay language, our study confirmed a known and complex relationship between food, stress and psychological factors when considering online forums as a complementary source of real-world evidence.

Conclusions

GI discomfort is an actively discussed topic in French web forums. When identified in a portion of active users, the gender and age of most active users tend to mirror the higher prevalence of DGBIs in women aged below 40 years. We were able to segment web users into several clusters corresponding to specific GI symptoms or diagnosed disorders and characterized by distinct demographic parameters and expression of variables related to stress. The main factors of GI discomfort as perceived by web users are food and psychological factors. This paper could benefit from a similar analysis based on additional sources to cover further languages (eg, English) to study the similarities and differences of the results at a larger scale and with different dietary and cultural backgrounds. To conclude, this innovative infodemiology approach has shown that identifying discussion topics associated with GI discomfort online is feasible and can serve as a complementary source of real-world evidence.

Multimedia Appendix 1

Corpus description.
[[DOCX File , 21 KB - jmir_v22i11e17247_app1.docx]

Multimedia Appendix 2
Number of extracted messages and associated number of web users per data source.
[[DOCX File , 22 KB - jmir_v22i11e17247_app2.docx]

Multimedia Appendix 3
List of keywords used for the extraction of messages.
[[DOCX File , 23 KB - jmir_v22i11e17247_app3.docx]

Multimedia Appendix 4
List of features used for the segmentation of web users.
[[DOCX File , 21 KB - jmir_v22i11e17247_app4.docx]

Multimedia Appendix 5
List of anatomical regions and pathophysiological mechanisms used for the segmentation of users.
[[DOCX File , 20 KB - jmir_v22i11e17247_app5.docx]

Multimedia Appendix 6
List of causality terms used for the identification of perceived factors.
[[DOCX File , 28 KB - jmir_v22i11e17247_app6.docx]

References

1. Jiang X, Locke GR, Choung RS, Zinsmeister AR, Schleck CD, Talley NJ. Prevalence and risk factors for abdominal bloating and visible distention: a population-based study. Gut 2008 Jun;57(6):756-763 [FREE Full text] [doi: 10.1136/gut.2007.142810] [Medline: 18477677]

2. Tielemans MM, Jaspers Focks J, van Rossum LG, Eikendal T, Jansen JB, Laheij RJ, et al. Gastrointestinal symptoms are still prevalent and negatively impact health-related quality of life: a large cross-sectional population based study in The Netherlands. PLoS One 2013;8(7):e69876 [FREE Full text] [doi: 10.1371/journal.pone.0069876] [Medline: 23922836]

3. Schmulson MJ, Drossman DA. What is new in Rome IV. J Neurogastroenterol Motil 2017 Apr 30;23(2):151-163 [FREE Full text] [doi: 10.5056/jnm16214] [Medline: 28274109]

4. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. Clin Gastroenterol Hepatol 2012 Jul;10(7):712-21.e4. [doi: 10.1016/j.cgh.2012.02.029] [Medline: 22426087]

5. Sperber AD, Dumitrascu D, Fukudo S, Gerson C, Ghoshal UC, Gwee KA, et al. The global prevalence of IBS in adults remains elusive due to the heterogeneity of studies: a Rome Foundation working team literature review. Gut 2017 Jun;66(6):1075-1082. [doi: 10.1136/gutjnl-2015-311240] [Medline: 26818616]

6. Pen CL, Ruszniewski P, Gaudin A, Amouretti M, Bommelaer G, Frexinos J, et al. The burden cost of French patients suffering from irritable bowel syndrome. Scand J Gastroenterol 2004 Apr;39(4):336-343. [doi: 10.1080/00365520310008458] [Medline: 15125465]

7. Frändemark Å, Törnblom H, Jakobsson S, Simrén M. Work productivity and activity impairment in irritable bowel syndrome (IBS): a multifaceted problem. Am J Gastroenterol 2018 Oct;113(10):1540-1549. [doi: 10.1038/s41395-018-0262-x] [Medline: 30254230]

8. Canavan C, West J, Card T. Review article: the economic impact of the irritable bowel syndrome. Aliment Pharmacol Ther 2014 Nov;40(9):1023-1034 [FREE Full text] [doi: 10.1111/apt.12938] [Medline: 25199904]

9. Cogliandro R, Antonucci A, de Giorgio R, Barbara G, Cremon C, Cogliandro L, et al. Patient-reported outcomes and gut dysmotility in functional gastrointestinal disorders. Neurogastroenterol Motil 2011 Dec;23(12):1084-1091. [doi: 10.1111/j.1365-2982.2011.01783.x] [Medline: 21917083]

10. Posserud I, Syrous A, Lindström L, Tack J, Abrahamsson H, Simrén M. Altered rectal perception in irritable bowel syndrome is associated with symptom severity. Gastroenterology 2007 Oct;133(4):1113-1123. [doi: 10.1053/j.gastro.2007.07.024]

11. Bennet SM, Polster A, Törnblom H, Isaksson S, Capronnier S, Tessier A, et al. Global cytokine profiles and association with clinical characteristics in patients with irritable bowel syndrome. Am J Gastroenterol 2016 Aug;111(8):1165-1176. [doi: 10.1038/ajg.2016.223] [Medline: 27272011]

12. Dunlop SP, Hebden J, Campbell E, Naesdal J, Olbe L, Perkins AC, et al. Abnormal intestinal permeability in subgroups of diarrhea-predominant irritable bowel syndromes. Am J Gastroenterol 2006 Jun;101(6):1288-1294. [doi: 10.1111/j.1572-0241.2006.00672.x] [Medline: 16771951]

XSL·FO
RenderX

13. Tap J, Derrien M, Törnblom H, Brazeilles R, Cools-Portier S, Doré J, et al. Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome. Gastroenterology 2017 Jan;152(1):111-23.e8 [FREE Full text] [doi: 10.1053/j.gastro.2016.09.049] [Medline: 27725146]

14. Zvolensky M, Jardin C, Farris SG, Kauffman B, Bakhshaie J, Garey L, et al. Gut interpretations: how difficulties in emotion regulation may help explain the relation of visceral sensitivity with depression and anxiety among young adults with gastrointestinal symptoms. Psychol Health Med 2018 Aug;23(7):840-845. [doi: 10.1080/13548506.2018.1455984] [Medline: 29580068]

15. Zamani M, Alizadeh-Tabari S, Zamani V. Systematic review with meta-analysis: the prevalence of anxiety and depression in patients with irritable bowel syndrome. Aliment Pharmacol Ther 2019 Jul;50(2):132-143. [doi: 10.1111/apt.15325] [Medline: 31157418]

16. Enck P, Koehler U, Weigmann H, Mueller-Lissner S. Abdominal pain, cramping or discomfort impairs quality of life in women: an internet-based observational pilot study focussing on impact of treatment. Z Gastroenterol 2017 Mar;55(3):260-266. [doi: 10.1055/s-0043-100022] [Medline: 28288501]

17. Brun-Strang C, Dapoigny M, Lafuma A, Wainsten JP, Fagnani F. Irritable bowel syndrome in France: quality of life, medical management, and costs: the Encoli study. Eur J Gastroenterol Hepatol 2007 Dec;19(12):1097-1103. [doi: 10.1097/MEG.0b013e3282f1621b] [Medline: 17998835]

18. Ford AC, Lacy BE, Talley NJ. Irritable bowel syndrome. N Engl J Med 2017 Jun 29;376(26):2566-2578. [doi: 10.1056/NEJMra1607547] [Medline: 28657875]

19. Mahon J, Lifschitz C, Ludwig T, Thapar N, Glanville J, Miqdady M, et al. The costs of functional gastrointestinal disorders and related signs and symptoms in infants: a systematic literature review and cost calculation for England. Br Med J Open 2017 Nov 14;7(11):e015594. [doi: 10.1136/bmjopen-2016-015594] [Medline: 29138194]

20. El-Salhy M, Ostgaard H, Gundersen D, Hatlebakk JG, Hausken T. The role of diet in the pathogenesis and management of irritable bowel syndrome (Review). Int J Mol Med 2012 May;29(5):723-731. [doi: 10.3892/ijmm.2012.926] [Medline: 22366773]

21. Monsbakken KW, Vandvik PO, Farup PG. Perceived food intolerance in subjects with irritable bowel syndrome- etiology, prevalence and consequences. Eur J Clin Nutr 2006 May;60(5):667-672 [FREE Full text] [doi: 10.1038/sj.ejcn.1602367] [Medline: 16391571]

22. McKenzie YA, Bowyer RK, Leach H, Gulia P, Horobin J, O'Sullivan NA, (IBS Dietetic Guideline Review Group on behalf of Gastroenterology Specialist Group of the British Dietetic Association). British dietetic association systematic review and evidence-based practice guidelines for the dietary management of irritable bowel syndrome in adults (2016 update). J Hum Nutr Diet 2016 Oct;29(5):549-575. [doi: 10.1111/jhn.12385] [Medline: 27272325]

23. Halmos EP, Gibson PR. Controversies and reality of the FODMAP diet for patients with irritable bowel syndrome. J Gastroenterol Hepatol 2019 Jul;34(7):1134-1142. [doi: 10.1111/jgh.14650] [Medline: 30945376]

24. Whitfield KL, Shulman RJ. Treatment options for functional gastrointestinal disorders: from empiric to complementary approaches. Pediatr Ann 2009 May;38(5):288-90, 92 [FREE Full text] [Medline: 19476303]

25. Fox S, Duggan M. Health Online. Pew Research Center. 2013 Jan 15. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-02-27]

26. Khan A, Baharudin B, Lee LH, Khan K. A review of machine learning algorithms for text-documents classification. JAIT 2010 Feb 1;1(1):-. [doi: 10.4304/jait.1.1.4-20]

27. Speier W, Ong MK, Arnold CW. Using phrases and document metadata to improve topic modeling of clinical reports. J Biomed Inform 2016 Jun;61:260-266 [FREE Full text] [doi: 10.1016/j.jbi.2016.04.005] [Medline: 27109931]

28. Zheng C, Rashid N, Wu Y, Koblick R, Lin AT, Levy GD, et al. Using natural language processing and machine learning to identify gout flares from electronic clinical notes. Arthritis Care Res (Hoboken) 2014 Nov;66(11):1740-1748 [FREE Full text] [doi: 10.1002/acr.22324] [Medline: 24664671]

29. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013 Mar;22(3):256-262 [FREE Full text] [doi: 10.1002/pds.3365] [Medline: 23322591]

30. Chen X, Faviez C, Schuck S, Lillo-Le-Louët A, Texier N, Dahamna B, et al. Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate. Front Pharmacol 2018;9:541 [FREE Full text] [doi: 10.3389/fphar.2018.00541] [Medline: 29881351]

31. Foulquié P, Voillot P, Faviez CM, Mebarki A, Chen X, Texier N, et al. Étude de l'usage du méthylphénidate sur les réseaux sociaux. Revue d'Épidémiologie et de Santé Publique 2018 Jun;66:S225 [FREE Full text] [doi: 10.1016/j.respe.2018.04.030]

32. Schück S, Foulquié P, Mebarki A, Voillot P, Texier N, Faviez C. Détection automatique du mésusage des neuroleptiques dans le trouble anxieux et la démence à partir des réseaux sociaux. Revue d'Épidémiologie et de Santé Publique 2018 Jun;66:S211-S212 [FREE Full text] [doi: 10.1016/j.respe.2018.04.048]

33. Schück S, Voillot P, Foulquié P, Faviez C, Mebarki A, Texier N, et al. Que nous apportent les réseaux sociaux quant à la crise sanitaire du levothyrox d'août 2017. Revue d'Épidémiologie et de Santé Publique 2018 Jun:s225 [FREE Full text] [doi: 10.1016/j.respe.2018.04.029]

34. Fagherazzi G, Ravaud P. Digital diabetes: perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]

35. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017 Jun 6;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]

36. Doctissimo Forums. 2020. URL: http://archive.today/SCVBL [accessed 2020-10-02]

37. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? J Med Internet Res Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]

38. Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-based signal detection using medical forums data in france: comparative analysis. J Med Internet Res 2018 Nov 20;20(11):e10466 [FREE Full text] [doi: 10.2196/10466] [Medline: 30459145]

39. Najork M. Web Crawler Architecture. Encyclopedia of Database Systems. URL: https://doi.org/10.1007/978-0-387-39940-9_457 [accessed 2020-10-02]

40. Ghosh DD, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system. Cartogr Geogr Inf Sci 2013;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]

41. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]

42. Yesha R, Gangopadhyay A. A Method for Analyzing Health Behavior in Online Forums. In: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics. 2015 Presented at: BCB'15; September, 2015; Atlanta, Georgia. [doi: 10.1145/2808719.2812592]

43. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. J Med Internet Res Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]

44. Blei DM, Lafferty JD. Text Mining: Classification, Clustering, and Applications. Boca Raton, FL: CRC Press; 2020.

45. Blei DM, Lafferty JD. Proceedings of the 23rd International Conference on Machine Learning. 2006 Jun Presented at: ACM'06; June 25-29, 2006; University in Pittsburgh, Pennsylvania.

46. Roberts M, Stewart B, Tingley D. Stm: an r package for structural topic models. J Stat Soft 2019 Oct 31;91(2):1-40 [FREE Full text] [doi: 10.18637/jss.v091.i02]

47. Lovell RM, Ford AC. Effect of gender on prevalence of irritable bowel syndrome in the community: systematic review and meta-analysis. Am J Gastroenterol 2012 Jul;107(7):991-1000. [doi: 10.1038/ajg.2012.131] [Medline: 22613905]

48. Pinto-Sanchez MI, Verdu EF. Non-celiac gluten or wheat sensitivity: it's complicated!. Neurogastroenterol Motil 2018 Aug;30(8):e13392. [doi: 10.1111/nmo.13392] [Medline: 30854780]

49. McNaughton DT, Andreasson A, Ljótsson B, Beath AP, Hush JM, Talley NJ, et al. Effects of psychology and extragastrointestinal symptoms on health care use by subjects with and without irritable bowel syndrome. Clin Gastroenterol Hepatol 2020 Apr;18(4):847-54.e1. [doi: 10.1016/j.cgh.2019.07.019] [Medline: 31323378]

50. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. Adv Nutr 2015 Jan;6(1):5-18 [FREE Full text] [doi: 10.3945/an.114.007492] [Medline: 25593140]

51. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. J Med Internet Res Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]

52. Schäfer F, Jeanne JF. Evaluating the effects of food on health in a world of evolving operational challenges. Contemp Clin Trials Commun 2018 Dec;12:51-54 [FREE Full text] [doi: 10.1016/j.conctc.2018.09.001] [Medline: 30259002]

53. Khazen W, Jeanne JF, Demaretz L, Schäfer F, Fagherazzi G. Rethinking the use of mobile apps for dietary assessment in medical research. J Med Internet Res 2020 Jun 18;22(6):e15619 [FREE Full text] [doi: 10.2196/15619] [Medline: 32554383]

54. Yao CK, Gibson PR, Shepherd SJ. Design of clinical trials evaluating dietary interventions in patients with functional gastrointestinal disorders. Am J Gastroenterol 2013 May;108(5):748-758. [doi: 10.1038/ajg.2013.77] [Medline: 23609614]

55. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]

56. mySymptoms App. URL: https://skygazerlabs.com/wp/ [accessed 2020-10-02]

57. Clevers E, Törnblom H, Simrén M, Tack J, van Oudenhove L. Relations between food intake, psychological distress, and gastrointestinal symptoms: a diary study. United European Gastroenterol J 2019 Aug;7(7):965-973 [FREE Full text] [doi: 10.1177/2050640619839859] [Medline: 31428421]

58. Qu'est-ce que le SNDS ? Système National Des Données De Santé. 2020. URL: http://archive.today/7DDam%20on%2002Aug2019 [accessed 2020-10-02]

59. Sabaté JM, Rivière S, Jouet P, Gastaldi-Menager C, Fagot-Campagna A, Tuppin P. Healthcare use by 30,000 patients with irritable bowel syndrome (IBS) in France: a 5-year retrospective and one-year prospective national observational study.

BioMed Central Gastroenterol 2019 Jun 27;19(1):111 [FREE Full text] [doi: 10.1186/s12876-019-1031-z] [Medline: 31248366]

60. Gupta RS, Warren CM, Smith BM, Jiang J, Blumenstock JA, Davis MM, et al. Prevalence and severity of food allergies among US adults. J Am Med Assoc Netw Open 2019 Jan 4;2(1):e185630 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.5630] [Medline: 30646188]

Abbreviations

DGBI: disorders of gut-brain interaction
GERD: gastroesophageal reflux disease
GI: gastrointestinal
IBS: irritable bowel syndrome
LDA: latent Dirichlet allocation
MedDRA: Medical Dictionary for Regulatory Activities
TF-IDF: term frequency-inverse document frequency

XSL•FO
RenderX

Original Paper

COVID-19 and the Gendered Use of Emojis on Twitter: Infodemiology Study

Ahmed Al-Rawi[1], PhD; Maliha Siddiqi[1], MA; Rosemary Morgan[2], PhD; Nimisha Vandan[3], PhD; Julia Smith[1], PhD; Clare Wenham[4], PhD

[1]Simon Fraser University, Burnaby, BC, Canada

[2]John Hopkins University, Baltimore, MD, United States

[3]Hong Kong University, Hong Kong, China (Hong Kong)

[4]London School of Economics, London, United Kingdom

Corresponding Author:
Ahmed Al-Rawi, PhD
Simon Fraser University
Schrum Science Centre-K 9653
Burnaby, BC, V5A 1S6
Canada
Phone: 1 7787824419
Email: aalrawi@sfu.ca

Abstract

Background: The online discussion around the COVID-19 pandemic is multifaceted, and it is important to examine the different ways by which online users express themselves. Since emojis are used as effective vehicles to convey ideas and sentiments, they can offer important insight into the public's gendered discourses about the pandemic.

Objective: This study aims at exploring how people of different genders (eg, men, women, and sex and gender minorities) are discussed in relation to COVID-19 through the study of Twitter emojis.

Methods: We collected over 50 million tweets referencing the hashtags #Covid-19 and #Covid19 for a period of more than 2 months in early 2020. Using a mixed method, we extracted three data sets containing tweets that reference men, women, and sexual and gender minorities, and we then analyzed emoji use along each gender category. We identified five major themes in our analysis including morbidity fears, health concerns, employment and financial issues, praise for frontline workers, and unique gendered emoji use. The top 600 emojis were manually classified based on their sentiment, indicating how positive, negative, or neutral each emoji is and studying their use frequencies.

Results: The findings indicate that the majority of emojis are overwhelmingly positive in nature along the different genders, but sexual and gender minorities, and to a lesser extent women, are discussed more negatively than men. There were also many differences alongside discourses of men, women, and gender minorities when certain topics were discussed, such as death, financial and employment matters, gratitude, and health care, and several unique gendered emojis were used to express specific issues like community support.

Conclusions: Emoji research can shed light on the gendered impacts of COVID-19, offering researchers an important source of information on health crises as they happen in real time.

Introduction

Background

COVID-19 has changed the way we communicate and interact with others. In an effort to maintain physical distancing and stop the spread of the virus, a lot of communication has moved from face-to-face (F2F) to online platforms including Twitter, Facebook, and Instagram, with users sharing information, messages, opinions, and beliefs about COVID-19 through these platforms. One form of online communication is through the use of emojis, defined as "a visual representation of an emotion,

idea or symbolism. It conveys concern, humor, anger or sarcasm" [1]. Billions of emojis are sent every day in different social media platforms [2], indicating their widespread use and popularity. For instance, the face with tears emoji () was named word of the year in 2015 and is considered the most used emoji across all genders [3].

This study focuses on how people of different genders (eg, men, women, and sex and gender minorities) are discussed in relation to COVID-19 through Twitter emojis. It asks the following research question: what are the gendered types of sentiments and meanings expressed with emojis with regard to the COVID-19 pandemic and how does emoji use differ when associated with different genders? We argue that, by exploring how the experiences of men, women, and sex and gender minorities in relation to COVID-19 are discussed using emojis, we can understand how emoji use represents or perpetuates (often inequitable) gender norms, roles, and relations in response to COVID-19, as well as how the pandemic may be impacting gender differently. Thus, analysis of emojis can provide a new methodological approach for rapid gender analysis in crisis settings, complementing more traditional forms of gender analysis through surveys or interviews.

The gendered dimensions of COVID-19 are becoming increasingly apparent. Although gendered dimensions of outbreaks have been demonstrated previously [4-6], given the widespread nature of COVID-19, these are becoming even more visible. First, women represent approximately 70% of health care workers worldwide and, thus, are disproportionately on the frontline of this battle against COVID-19 [7]. This formal care role is extended into homes, where informal care norms dictate that women provide care to sick family members—thereby exposing themselves to risk of infection and assuming the burden of work associated with this additional care [4]. This care burden also extends to the additional work created by having all family members at home during lockdown and the requirements for homeschooling as facilities remain shut. Women are increasingly seen to do this informal care work on top of their routine paid employment, with recent data demonstrating that women perform these domestic roles on top of paid employment considerably more than men, even when they are the main breadwinner in a household [8]. However, women are also disproportionately affected by job losses and furlough schemes, with data demonstrating that more women have been made redundant or furloughed than men [9]. This raises concerns for women's economic empowerment and role in the labor force going forward. Women have also experienced increasing rates of domestic violence during lockdown, with calls to domestic violence hotlines surging in March and April 2020 when lockdown measures were enforced [10]. Finally, women's differential health needs have been affected by COVID-19, with access to sexual reproductive health services and maternity care limited, reduced, or cancelled and deemed nonessential as resources are diverted to COVID-19 [11]. These illustrative examples demonstrate the vital need for gender analysis during outbreaks and in as real time as possible to outline the emerging gendered needs to policy makers throughout the crisis and postcrisis period.

Using emojis to relay ideas about health or disease is not new. For example, unique emojis have been used and introduced in relation to pandemics like malaria and Zika to raise awareness about their risks [12]. During the COVID-19 crisis when a shift from in-person conversations to a virtual communication paradigm occurred, emojis became even more vital to public discourse. For example, women in Ecuador are sending coded messages through emojis asking for assistance to escape domestic violence [13]. We also found many health-related emojis used in relation to all three gender groups, such as hospitalization (), medical services (), emergency sign (), medicine (), and syringe (). In addition, new types of emojis have emerged to represent key messages during the pandemic, including *wear a mask* () and *microbes* (), and a sequence of emojis provide specific social guidelines such as *do not sneeze into your hand* () or *keep social distance* (). Another interesting aspect of emoji sequences is emphasis, as repeated use of one emoji like thank you () shows deeper appreciation than using the same emoji once. We know of no previous empirical research on the gender elements of health-related emoji use generally or on COVID-19 specifically.

The Conceptualization of Emojis

Since the immense popularization of social media, coupled with the technological advancements in smartphone technology, emojis have witnessed an increasingly widespread use among different age groups and genders. Stark and Crawford [14] argued that emojis act as historical, social, and cultural objects, forming a type of a language that can help to underscore tone and communicate humor, allowing users to express their personality through their online interactions and relay a form of *digital feeling*. Emojis also allow users to express the characteristics that inform their individual identities including gender, race, age, and disability or demonstrate ideas or objects that are important to them. As such, exploring the use of emojis can provide important cultural and historical understanding into how people communicate, express themselves, or disseminate normative ideas and beliefs, especially during a public health crisis like the COVID-19 pandemic. In this study, we regard emojis as embodiments of affective expressions and "cultural objects" [14].

Generally speaking, the terms emojis and emoticons are often used interchangeably, although technically they connote different meanings. Although both are used as supplemental devices to nonvisual communication, emojis are pictorial representations, while emoticons are combinations of letters and punctuation marks available on smartphone and computer keyboards. For the purpose of this study, we treated both terms as interchangeable while concentrating our research interest mostly on emoji use. Emoticons resemble facial nonverbal behavior and may serve at least some of the same functions as nonverbal behavior in F2F communication [1].

Writing down emotional messages changes the intensity of the emotion because there is time to read over the text and reflect on one's emotional state [15]. Emojis may enhance the exchange of emotional information by providing additional social cues

beyond those found in a text message [16] used to augment the meaning of a message as a whole [17]. Emojis are also seen less as messaging tropes and more as expressive devices. Despite these strong positives for the use of emojis, there remains a perception in some quarters that their use is a lesser form of language and is devaluing and devolving language. However, these concerns are not only unfounded, but they ignore human needs for nonverbal communicative practices [18].

Emojis convey various forms of sentiments and messages that vary across cultures that need further scholarly attention. For example, in a multiphasic big data analysis based upon more than a million tweets and using Geert Hofstede's national cultural scores, researchers established an unlikely relationship between cultures and vertical and horizontal emoticon use. In this regard, individualistic countries show a suppressed use of vertical emotions (emoticons emphasizing eye shape), whereas collectivistic cultures favor less horizontal emoticons (emoticons emphasizing mouth shape) [19]. Another more recent study reiterates the nonuniversality of emoticons through the analysis of selected populations of Tanzania, Cameroon, and Japan. Although Japanese people were sensitive to the different emotions embedded in emoticons, Cameroonian and Tanzanian people hardly read emotion from emoticons [20]. Further, Cheng [21] conducted an experiment that involved a study of sadness conveyed through both emoticons and emojis by Spanish and Chinese participants. The study confirmed that Spanish users prefer plain text messaging more than their Chinese counterparts, the latter registered a higher use of sad emojis and emoticons [21]. In brief, there are cultural differences in the way we use emojis and their types. Though it has not been empirically studied in previous research, many emojis are polysemous, as they can have more than one meaning, depending on cultural contexts and individual users.

The Gendered use of Emojis

Although a plethora of research exists on emoji use itself, there is scarce scholarly knowledge that explores how gender norms, roles, and relations are represented within and perpetuated by emojis; instead, the bulk of previous research seems to be mostly focused on cross-cultural use of emojis, as previously indicated. During the beginning of emoji production, women were portrayed in stereotypical representations in emoji libraries; most activity-based emojis representing women were either brides or dancers, or exuded seductress characteristics [22]. Although the variations of smiley face emoji represent neither men or women, emojis were not always so gender neutral, though they were considered so when first released. For example, *neutral* images such as of a doctor or police officer were used to represent particular professions. The original emojis for such professions, however, represented the male body and clearly depicted men in these professions, reinforcing gendered norms and biases related to who engages in these types of professions. As Caroline Criado Perez [23] reports in her book, *Invisible Women*, what was striking was that it was not the original code that delineated these emojis as male but the platforms that interpreted gender-neutral terms as male.

In general, women were restricted to cliché portrayals while exempted from role depictions such as surgeons, lawyers, and teachers. This bias remained until 2016 when the original code was redesigned to gender all emojis. Today, male and female options exist for all professions and athletes, which is important because they act as a mechanism to perpetuate or reinforce inequitable gender norms, roles, and relations, which have had and continue to have a negative impact on peoples' lives and health [24].

Most studies have shown varying results that women use emojis significantly more than men [25,26], and some scholars observe there are certain overlaps in the use patterns. Wolf [27], for instance, conducted research on a total of 251 posts on the USENET platform (a primitive form of data platforms used to read and distribute news posts) to examine behavioral patterns among women and men when it comes to emoji use. After determining the variety of the emoticons used, three main categories were established: smileys (positive), frowns (negative), and winks (sarcastic, funny, or flirting). In addition, the frequency of emoticon use was tabulated along gender use [27]. The results reiterated the idea of an emotional woman and an inexpressive man, and presented an interesting discussion about the blurring lines between the definitions of gender in emoji use, particularly the commonalities in employing humor by women and sarcasm by men [27]. In addition, Kalsoom and Kalsoom [22] used a semiotics approach and a feminist paradigm to map the meaning making potential of semiotic resources and critically examined stereotypical and professional women emojis. For example, the bunny girl emoji () often refers to the objectification of women through a *showgirl* representation with costumes and dancing, while professional emojis highlight different professional roles such as scientist and doctor along different genders [22]. Danesi [28] uses a similar semiotic yet nontechnical approach to understand meaning making through signs and symbols of emojis, which he terms as a rather generic tool. On the other hand, an extensively thorough study of blogs used by people between the ages of 13-19 years revealed that men and boys have a tendency toward using emoticons coupled with active and resolute language, while women and girls used them to express strong social interactions [29].

Further, Kavanagh [30] collected posts from American and Japanese blogs and found that women's emoticons were dominant in both high context culture (such as that of the Japanese), where communication is more indirect and symbolic, and low context cultures (such as America), where communication is more direct and succinct [30]. Further, Tossell et al [31] investigated how emoticons were used and, in particular, how gender differences exist in the frequency and variety of emoticons. For their analysis, data from 21 smartphones was taken over a 6-month period. In terms of quantity, the authors observed that women were more likely to use emoticons than men, while the latter preferred using a distinct range of emoticons to express themselves. Drawing on Tosell et al's [31] findings, Shahbaz et al [32] conducted a study on the users of Kika Keyboard, a major Google play application with a diverse library of 1281 emojis. The authors found that "there are stark differences in the emoji usage preferences in men and women with women using more than one emoji in a single message and men using them in consecutive turns."

Another research conducted in 2018 at the Peking University in China commented extensively on the ubiquitous nature of emoticon use and its ability to surpass language barriers and travel worldwide. The study distinguishes between expressions among women and men, and suggests designing new gender-based machine learning modules [33].

Finally, diversifying the gender demographic in terms of age, Nishimura [34] conducted a qualitative study with 50 Japanese bloggers older than 60 years and men and women in their twenties and thirties to understand their emoticon use. The outcomes showed that younger women are more active in using emotions compared to other genders and age groups, followed by older women and older men. Young men in the age group of 20-40 years showed the lowest use of emoticons.

What is obviously lacking in previous research on gendered emojis is a focus on public health issues; hence, our study fills a gap in literature. In addition, research on sexual and gender minorities' emoji use is still missing. This reveals a persistent pattern that excludes minority groups from scholarly representations of who is considered equal members within society. Buff [35] in her proposal to the Unicode consortium, which works to create universal international software standards, not only pushed for a much needed reform toward the introduction of gender-neutral signs but also the inclusion of the *third gender* in emojis, stressing the risks of solely highlighting gender binary models that can perpetuate harmful nonbinary gender stereotypes and outdated world views.

Our study attempts to explore how people of different genders are mentioned and discussed in connection to COVID-19 through the study of social media emojis.

Methods

Through Twitter, we collected 50,811,299 tweets referencing #Covid-19 and #Covid19 that were posted by 11,706,754 unique users. The tweets were collected for a period of over 2 months from February 12 until April 18, 2020. We believed that the large number of tweets was enough to conduct our study on gendered emoji use because the data set contained a wide range of emojis, but it was not empirically possible to ascertain the sexual and gender backgrounds of social media posters, which remains a limitation in our study. This data set was collected using the TCAT platform that uses Twitter public application programming interface (API), allowing about 1% of public tweets to be fetched. Due to API limitations, the platform often hits the rate limit allowed, so a brief delay sometimes happens in collecting tweets. In other words, we have not violated the terms and conditions of the Twitter platform in our data collection. We then used several customized Python programming packages to first extract three gender-specific tweets (men, women, and nonbinary) and then fetch emojis from each data set (Table 1) [36].

Table 1. The top 20 most frequent emojis along gender groups.

No	Women	Frequency, n	Men	Frequency, n	Nonbinary	Frequency, n	All data	Frequency, n
1		16,496		23,538		837		985,584
2		14,179		22,239		707		797,255
3		9732		20,709		546		720,706
4		9347		5894		451		619,156
5		9187		4338		344		595,090
6		8783		3933		302		570,762
7		6281		3078		302		568,700
8		5718		2740		280		473,298
9		5574		2723		255		468,919
10		5461		2602		226		456,706
11		5376		2460		221		452,459
12		5054		2163		217		449,328
13		5035		2018		212		404,043
14		4655		1900		197		399,166
15		4502		1775		188		397,510
16		4303		1416		184		367,933
17		4251		1408		180		342,812
18		4166		1219		161		337,518
19		3726		1137		156		296,252
20		3513		1091		147		293,830

Though the search terms were not exhaustive and only relied on the English language, the search terms used to extract tweets on women were "woman*," "women," "femin*," "girl*," "lady," "ladies," and "female*," and the total number of English language tweets referencing the aforementioned search terms was 541,698 tweets, constituting 1% of the total data set of 50,811,299 tweets. These filtered tweets were sent by 367,037 unique users. As for men, the search terms used were "boy*," "man*," "men," "gentlem*," and "mascul*," and the total number of English language tweets referencing the aforementioned search terms was 297,155 tweets sent by 231,899 unique users. To gather data related to nonbinary genders and those often marginalized within binary gender constructs, we used the search terms "nonbinary," "non-binary," "trans," "transgender," "two spirited," "two-spirited," "LGBT*," "gay*," "homosexual*," "lesbian," "bisexual*," and "queer," which were classified under the heading "sexual and gender minorities" in the analysis that follows, which we recognized combines a vast range of gender and sexual identities. The total number of English language tweets referencing the aforementioned search terms was 26,048 posted by 20,744 unique users. These search terms were agreed upon after consultation with the research team members, and we recognize that they are not exhaustive, as gender identities and corresponding terms are constantly changing across various cultures.

Since the total number of extracted emojis were large in number (n=33,705,203) with 1297 different types, we focused on the top 600 emojis along the three data sets (the top 200 emojis representing each gender), using a manual approach in coding the emojis. We used a similar approach to Shahbaz et al [32] and Chen et al's [33] studies, which focused on gender-based analysis of emoji use patterns including frequency, choice, and consecutive and discrete patterns. We believe that some of these patterns, especially the frequency of emojis, shed important light on the salience of certain public sentiments expressed on

XSL•FO

RenderX

Twitter because the frequency indicates the salience given to some emojis or sentiments. For example, the significance of using one emoji thousands of times should be different from another emoji used only a couple of times.

The other strength of this approach to quantifying emojis is that it illustrates significant differences in the number of emojis used to represent the different identified genders (see Table 1), so it is more accurate to take into account when calculating the percentages of each gender. Though there are several automated methods to code emojis into their assigned sentiments [37], our study uses a combined novel approach of semiotic analysis to understand the connotative and denotative emoji expressions [22,28] and manual sentiment analysis [38].

To classify emojis into their assigned sentiments, we used the database of emoji sentiment rankings, which is based on Novak et al's [39] emoji study. Their study investigated 1.6 million multilingual tweets (in 13 European languages), and human annotators manually classified and coded 751 different emojis found in these tweets by classifying them based on their sentiment ranking, indicating how positive, negative, or neutral each emoji was. In our study, we used negative, positive, and neutral values in measuring the sentiments of emojis instead of employing a Likert scale approach. This is because of the difficulty of obtaining intercoder reliability with a scalar approach, especially because the emojis we examined were uniquely related to the COVID-19 pandemic. For example, many emojis in our data sets were not listed in the aforementioned database. As a result, we used two other sources.

The first one was Unicode.org that lists more than 3300 emojis [40], and the second one is called emojipedia [41]. The latter is a consolidated database of most available emojis on modern smart devices, and it lists the varied meanings and interpretations of each emoji. These emoji repositories have been used in previous studies dealing with a variety of topics [42,43]. To make sure that our approach was valid and before we conducted the full scale study, two coders independently examined a representative 10% sample of the data sets (n=60), and the agreement between them was α≥.815 using Krippendorff alpha [44]. Finally, another novel approach that we followed in this study was examining some of the relevant emoji sequences because a combination of some emojis can denote a clearer message in the context of the pandemic, such as (), which means "if you are sick, wear a mask," or (), which denotes

that "coronavirus has made many people sick around the world." Other sequences include (), which refers to "please wash your hands before eating," or (), indicating "thank you or please wear a mask as an expression of solidarity." To identify these emoji sequences, we used a customized Python script and searched in the three data sets.[36]

Results

Data

This study is focused on examining the gendered public discourses using emojis. To answer the study's main research question, we found that gendered sentiments around COVID-19 were overwhelmingly positive alongside all genders, such as the predominant use of the *thank you* () emoji (see Figure 1). Emojis used alongside discourses around men, however, showed a significant higher tendency toward using positive sentiments (112,516/140,056, 80.3%) as compared to the other two gendered groups: women (174,741/243,302, 71.8%) and sexual and gender minorities (8186/11,849, 69%). Yet emojis used alongside discourses around men and women showed similar trends when it came to the use of negatively ranked emojis (men: 13,476/140,056, 9.6%; women: 22,167/243,302, 9.1%) such as the *heartbreak* () emoji compared to emojis used alongside discourses around sexual and gender minorities that showed the highest percentage of negativity (1315/11,849, 11%). As for neutral emojis, which neither indicate clear positive or negative sentiments like the neutral face () emoji, the analysis showed that discourses around sexual and gender minorities received the highest percentage (2348/11,849, 19.8%), followed by discourses on women (46,394/243,302, 19%), while the lowest use was found around discourses on men (14,064/140,056, 10%). In the following section, we present the five major themes that we qualitatively identified following the semiotic approach, which include morbidity and health fears, economic and employment concerns, health responses, support for frontline workers, and unique gendered emojis.

In the twitter data set, the total sum of the emoji counts (total number of times that emojis appeared) were as follows: 162,423 for men; 291,838 emojis referencing women; and 13,985 emojis mentioning sexual and gender minorities. The major themes that emerged from studying emojis based on their frequencies and percentages are discussed in the next sections.

Figure 1. The percentage of emoji sentiments along the three main genders. This figure was generated using Tableau and Photoshop.
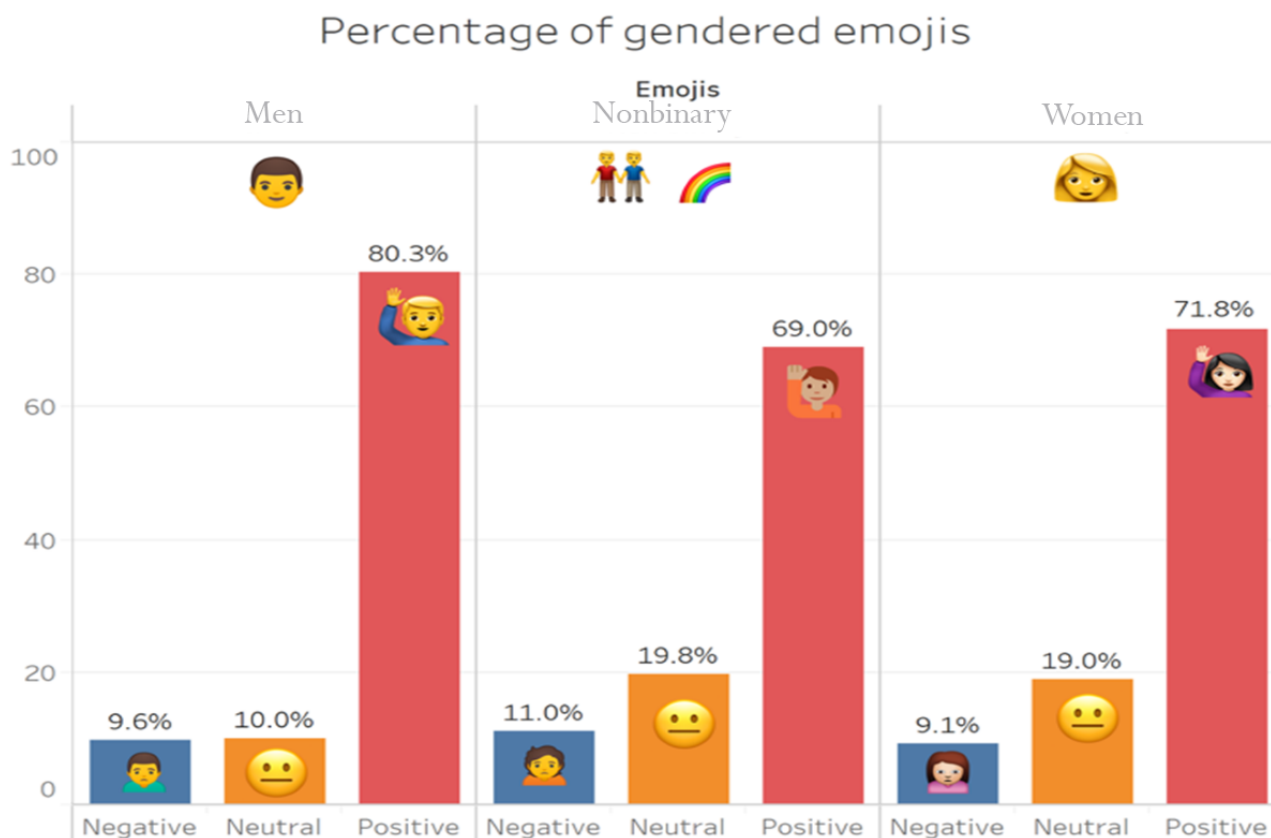
Morbidity and Health Fears

Twitter users employ different emojis to express fear, risk, and concerns that are noticeable in relation to the novel coronavirus, such as the coffin ([×]) and skull ([×]) emojis, in reference to death, as well as the ambulance siren ([×]) emoji, which often refers to the emergency siren vehicles carrying patients with COVID-19. In terms of gender differences, the coffin emoji was only found alongside discourses on men (n=129). This may be a reflection of how men and women are differentially perceiving the severity of COVID-19 due to sex differences in mortality among males and females, for male patients (in most contexts) are more likely to die due to COVID-19 compared to female patients [45]. In addition, the skull emoji was also mostly found in the discourses used around men (197/162,453, 0.12%) and, to a lesser extent, in relation to the discourses used around sexual and gender minorities (6/13,985, 0.04%). The siren emoji was used in the discourses around men (23,538/162,453, 14.7%) more than around women (2310/291,838, 0.8%) and sexual and gender minorities (56/13,985, 0.4%). Finally, the emoji sequence ([×]) was only found in the men's data set (n=37), and the emergency emoji sequence ([×]) was the 12th most frequent sequence in the men's data set. Another unique sequence in discussions about men was the following ([×]), which is a warning that not wearing a face mask can lead to death.

Economic and Employment Concerns

Another type of concern expressed online was related to financial matters. As the world experienced an economic crisis with uncounted layoffs and high unemployment rates during the pandemic, several emojis were used like money ([×]) and bank ([×]) to indicate economic instability, lack of employment opportunities, limited access to money, or financial concerns. Here, we found that the money emoji was not used in the discourses around men, and it was present in the discussions related to sexual and gender minorities (19/13,985, 0.13%) and in those related to women (1868/291,838, 0.6%), yet the bank emoji was only found in the discourses around women (n=4251, 1.5%).

As a significant transition was made from in-person to online operations that entailed a work-from-home dynamic, we found that the emoji that portrayed this phenomenon was the PC ([×]), which was again mostly used in relation to discourses around women (n=14,179, 15%), and much less in relation to discourses around sexual and gender minorities (n=18, 0.16%) and men (n=88). Another relevant emoji that was only found in the discourses on women was the *working from home* emoji ([×]; 496/291,838, 0.17%). The large difference in the use of these two latter emojis reflects the public discourses on the opportunities and challenges that women are facing during the pandemic. In Canada, for instance, women are more likely than men to hold jobs that can be performed from home during the pandemic [46]. These gender disparities shed light on the prominence of this emoji in the discourses on women.

Health Responses

We also examined how online communities responded toward COVID-19 in relation to health considerations. For example,

emojis were used to express health concerns like vomiting (), sneezing (), and fever (), in addition to items that can be used to protect oneself from contracting COVID-19, such as the soap bar emoji () and face mask emoji (). In relation to the gender differences, the soap emoji was not used in the discourses on men, but was found instead in the discourses around women (723/291,838, 0.2%) and, to a much lesser extent, discourses around sexual and gender minorities (8/13,985, 0.02%), while the face mask emoji was mostly used within discourses around women (3513/291,838, 1.2%), followed by men (660/162,453, 0.4%) and sexual and gender minorities (41/13,985, 0.2%).

In terms of the emoji sequence, we found that the women's data set contained more health instructions and guidelines than the other two data sets. For example, the following two emojis (; n=273) and (; n=19) refer to washing with water and soap, which can help stop the virus, while () refers to "women's testing" (n=59), () refers to "thank you or please wear a facemask" (n=42), and () refers to "thank you or please stay at home" (n=22). Other sequences were longer, such as (), which indicates "testing, wearing a mask, and taking cautious steps can stop the virus" (n=245). Although they were much less frequent, the following emoji sequences were found in the men's data set: (), which signifies the need to wear a mask if you go outside, and () and (), each showing solidarity for wearing the mask. Interestingly, the last two emoji sequences represent people from different racial backgrounds based on the skin tones used.

Support for Frontline Workers

There was an interesting positive trend manifested in the emojis that showed support and appreciation for frontline health care workers, including the rainbow emoji () that is used to show solidarity with the lesbian, gay, bisexual, transgender, and queer community (see Table 1) but has also been recently employed to denote unity, togetherness, as well as good fortune during the COVID-19 pandemic. The use of this symbol to support frontline workers, however, has also created some concerns among sexual and gender minorities due to their need to use the flag to remind people of their causes and struggle to achieve equality [47]. Other positive emojis that express support for frontline workers included the applause emoji (), the folding hands as a thankful gesture emoji (), and the globe or global solidarity () emoji. Here, we found some interesting gender differences in the thankful gesture, which was mostly used in discourses around women (16,496/291,838, 5.6%) compared to men (2602/162,423, 1.6%) and sexual and gender minorities (212/13,985, 1.5%). However, the applause emoji was used in similar percentages along the different genders (women: 4655/291,838, 1.6%; men: 2460/162,423, 1.5%; sexual and gender minorities: 221/13,985, 1.6%). Regarding the emoji frequencies, we found expressions of gratitude for hospital

workers such as (; n=42) and (; n=22) in the women's data set.

Unique Gendered Emojis

Finally, gender-specific emojis were often used in relation to women. For example, women were often represented through exercise such as doing yoga (), weight lifting (), or running (). These emojis were not found in the top emojis examined in the study in the discourses around men or sexual and gender minorities, possibly indicating a higher concern and attention paid by women to maintain physical health during the pandemic. In addition, there were many public discussions related to pregnant women (; n=1441), breastfeeding (; n=450), and a milk bottle (; n=418), representing the unique challenges women face due to their reproductive role. Regarding emoji sequences, we found several combinations referring to the aforementioned activities, such as (; n=23), in the women's data set.

Discussion

Principal Findings

Based on the previously mentioned findings on Twitter emojis, we found different discourses around men, women, and gender and sexual minorities. In general, emojis were positively associated with men while more negatively connected to sexual and gender minorities and women. In addition, there were unique emojis associated with women and the issue of health care workers as well as hygiene and with men and the issue of mortality.

Since the Twitter data was collected from different countries around the world, it is not possible to ascribe the aforementioned results to one country or culture, but we did find some familiar findings. For instance, there was a clear tendency to discuss men through positive sentiments, unlike the case of sexual and gender minorities, and to a lesser extent women, which might be related to the male-dominated Twitter platform that is estimated to be used by 62% men versus 38% women [48], as well as patriarchal cultural norms that span geographies. What is particularly notable here is the tendency to use negative emojis when discussing sexual and gender minorities, as this negativity appears to reflect ongoing stigma and discrimination toward these groups throughout recent history [49] and the way, for example, gay men were negatively affected by state policies in North America during the AIDS epidemic in the 1980s [50].

In addition, emojis offer insight into the COVID-19 impact on different genders. For example, we found a higher number of coffin and skull emojis when men were discussed, and generally, men have a higher mortality rate than women due to COVID-19. In addition, the financial burdens and concerns experienced by women, as well as the experiences of working from home, were reflected in the higher use of these emojis in relation to women. Interestingly, there is a prevalence in using the hand washing emojis in discussions about women, and there is evidence that responsibility for personal and family health and hygiene is

most often born by women and that men are less likely to wash their hands or wear a face mask [51]. The more dominant use of emojis such as praying hands in relation to women may show appreciation for these health-related roles. As a matter of fact, the majority of frontline workers worldwide are women [52].

These findings indicate that the analysis of emoji use offers an innovative methodological approach to analyzing the gendered impacts of health crises like COVID-19. This is particularly important as there are frequent calls for rapid gender analysis of health emergencies [53], and research through more established means such as surveys and interviews can be slow and completed after the fact. Emojis present a readily available, real-time data source that depicts discourses on gendered impacts of health crises as they are happening. However, this is the first study, which we know of, to employ emoji analysis in this way. There is a need to further develop this method and expand the analysis in terms of both depth and scope.

Conclusion

This study examined the gendered use of emojis on Twitter in relation to COVID-19, and the findings showed many differences alongside discourses of men, women, and gender minorities when certain topics were discussed, such as death, financial and employment matters, gratitude, and health care. In addition, there were several unique gendered emojis that were used to express specific issues like community support. In general, emojis are positively associated with men while more negatively connected to sexual and gender minorities and, to a lesser extent, women.

Our study is limited to the empirical examination of emoji use on Twitter in the context of COVID-19. It would be interesting to compare the emojified gendered public discourses on other platforms like Instagram and Facebook. Though it was not prominent in our top posts, emojis denoting race and ethnicity need to be incorporated as another important variable to be examined, and it would be interesting to conduct comparative studies on this aspect of emoji use. Subsequent studies might provide greater disaggregation of data related to sex and gender minority groups. Finally, our research is similar to previous studies on sentiment analysis that have limitations in terms of the dictionaries or word lists used [54] and the lack of examining the nature of users interacting online [55].

Conflicts of Interest

None declared.

References

1. Derks D, Bos AER, von Grumbkow J. Emoticons and online message interpretation. Soc Sci Comput Rev 2007 Dec 03;26(3):379-388. [doi: 10.1177/0894439307311611]
2. Statistics. World Emoji Day. 2020. URL: https://worldemojiday.com/statistics/ [accessed 2020-08-31]
3. Wirza Y, Nurbaeti DH, Hanifah H, Hanifah H. The difference in emoji usage between genders. In: Proceedings of the Twelfth Conference on Applied Linguistics (CONAPLIN 2019). 2019 Presented at: Twelfth Conference on Applied Linguistics (CONAPLIN 2019); 2019; Bandung, Indonesia. [doi: 10.2991/assehr.k.200406.047]
4. Harman S. Ebola, gender and conspicuously invisible women in global health governance. Third World Q 2016 Mar 10;37(3):524-541. [doi: 10.1080/01436597.2015.1108827]
5. Smith J. Overcoming the 'tyranny of the urgent': integrating gender into disease outbreak preparedness and response. Gender Dev 2019 Jun 28;27(2):355-369. [doi: 10.1080/13552074.2019.1615288]
6. Davies SE, Bennett B. A gendered human rights analysis of Ebola and Zika: locating gender in global health emergencies. Int Aff 2016 Aug 31;92(5):1041-1060. [doi: 10.1111/1468-2346.12704]
7. Boniol M, McIsaac M, Xu L, Wuliji T, Diallo K, Campbell J. Gender equity in the health workforce: analysis of 104 countries. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311314/WHO-HIS-HWF-Gender-WP1-2019.1-eng.pdf?ua=1
8. Schulte B, Swenson H. An unexpected upside to lockdown: men have discovered housework. The Guardian. 2020 Jun 17. URL: https://www.theguardian.com/us-news/2020/jun/17/gender-roles-parenting-housework-coronavirus-pandemic [accessed 2020-08-31]
9. COVID Inequality Project. Google Sites. 2020. URL: https://sites.google.com/view/covidinequality/
10. Taub A. A new COVID-19 crisis: domestic abuse rises worldwide. The New York Times. 2020 Apr 06. URL: https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html [accessed 2020-08-31]
11. Church K, Gassner J, Elliott M. Reproductive health under COVID-19 - challenges of responding in a global crisis. Sex Reprod Health Matters 2020 Dec;28(1):1-3. [doi: 10.1080/26410397.2020.1773163] [Medline: 32441213]
12. Pardes A. The complete history of emoji. Wired. 2020. URL: https://www.wired.com/story/guide-emoji/ [accessed 2020-08-31]
13. Rigby J. Women in Ecuador sending emoji coded pleas for help as domestic violence 'skyrockets' in lockdown. The Telegraph. 2020 Jun 18. URL: https://www.telegraph.co.uk/global-health/women-and-girls/women-ecuador-sending-emoji-coded-pleas-help-domestic-violence/ [accessed 2020-08-31]
14. Stark L, Crawford K. The conservatism of emoji: work, affect, and communication. Soc Media Soc 2015 Oct 08;1(2):205630511560485. [doi: 10.1177/2056305115604853]

15.   Fischer A. Gendered social interactions in face-to-face and computer-mediated communication. In: Kappas A, Krämer NC, editors. Face-to-Face Communication Over the Internet: Emotions in a Web of Culture, Language, and Technology. Cambridge, UK: Cambridge University Press; 2011.

16.   Thompsen PA, Foulger DA. Effects of pictographs and quoting on flaming in electronic mail. Comput Hum Behav 1996 Jun;12(2):225-243. [doi: 10.1016/0747-5632(96)00004-0]

17.   Rezabek L, Cochenour J. Visual cues in computer-mediated communication: supplementing text with emoticons. J Vis Literacy 2016 Feb 29;18(2):201-215. [doi: 10.1080/23796529.1998.11674539]

18.   Alshenqeeti H. Are emojis creating a new or old visual language for new generations? A socio-semiotic study. Adv Lang Literary Stud 2016 Dec 07;7(6):56-69. [doi: 10.7575/aiac.alls.v.7n.6p.56]

19.   Park J, Baek YM, Cha M. Cross-cultural comparison of nonverbal cues in emoticons on Twitter: evidence from big data analysis. J Commun 2014 Mar 19;64(2):333-354. [doi: 10.1111/jcom.12086]

20.   Takahashi K, Oishi T, Shimada M. Is ☺ smiling? cross-cultural study on recognition of emoticon's emotion. J Cross-Cultural Psychol 2017 Oct 24;48(10):1578-1586. [doi: 10.1177/0022022117734372]

21.   Cheng L. Do I mean what I say and say what I mean? A cross cultural approach to the use of emoticons and emojis in CMC messages. Fonseca, J Commun 2017;15(15):199. [doi: 10.14201/fjc201715199217]

22.   Kalsoom N, Kalsoom S. Semiotic representation of gender in Google emojis: a liberal feminist perspective. Linguistics Lit Rev 2019 Oct;05(02):127-137. [doi: 10.32350/llr.52.04]

23.   Criado-Perez C. Invisible Women: Data Bias in a World Designed for Men. New York City, NY: Abrams Press; 2019.

24.   Heise L, Greene ME, Opper N, Stavropoulou M, Harper C, Nascimento M, et al. Gender inequality and restrictive gender norms: framing the challenges to health. Lancet 2019 Jun;393(10189):2440-2454. [doi: 10.1016/s0140-6736(19)30652-x]

25.   Hwang HS. Gender differences in emoticon use on mobile text messaging: evidence from a Korean sample. Int J Journalism Mass Comm 2014 Dec 17;1(1):107. [doi: 10.15344/2349-2635/2014/107]

26.   Prada M, Rodrigues DL, Garrido MV, Lopes D, Cavalheiro B, Gaspar R. Motives, frequency and attitudes toward emoji and emoticon use. Telematics Inform 2018 Oct;35(7):1925-1934. [doi: 10.1016/j.tele.2018.06.005]

27.   Wolf A. Emotional expression online: gender differences in emoticon use. CyberPsychology Behav 2000 Oct;3(5):827-833. [doi: 10.1089/10949310050191809]

28.   Danesi M. The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet. London, England: Bloomsbury Academic; 2017.

29.   Huffaker DA, Calvert SL. Gender, identity, and language use in teenage blogs. J Computer-Mediated Commun 2006;10(2):00-00. [doi: 10.1111/j.1083-6101.2005.tb00238.x]

30.   Kavanagh B. A cross-cultural analysis of Japanese and English non-verbal online communication: the use of emoticons in weblogs. Intercultural Commun Stud 2010:3.

31.   Tossell CC, Kortum P, Shepard C, Barg-Walkow LH, Rahmati A, Zhong L. A longitudinal study of emoticon use in text messaging from smartphones. Comput Hum Behav 2012 Mar;28(2):659-663. [doi: 10.1016/j.chb.2011.11.012]

32.   Shahbaz M, Razzaq T, Gohar MR. Emoji usage by Android consumers: a gender-based inquiry. Orient Res J Soc Sci 2017:1-26.

33.   Chen Z, Lu X, Ai W, Li H, Mei Q, Liu X. Through a gender lens: learning usage patterns of emojis from large-scale Android users. In: WWW '18: Proceedings of the 2018 World Wide Web Conference. 2018 Presented at: WWW '18: The Web Conference 2018; April 2018; Lyon, France p. 763-772. [doi: 10.1145/3178876.3186157]

34.   Nishimura Y. A sociolinguistic analysis of emoticon usage in Japanese blongs: variation by age, gender, and topic. AoIR Selected Papers Internet Res 2015:17.

35.   Proposal for fully gender-inclusive emoji. Unicode. 2017. URL: https://www.unicode.org/L2/L2017/17232-gender-gap.pdf

36.   Al-Rawi A, Shukla V. Bots as Active News Promoters: A Digital Analysis of COVID-19 Tweets. Information 2020 Sep 27;11(10):461. [doi: 10.3390/info11100461]

37.   Fernández-Gavilanes M, Juncal-Martínez J, García-Méndez S, Costa-Montenegro E, González-Castaño FJ. Creating emoji lexica from unsupervised sentiment analysis of their descriptions. Expert Syst Applications 2018 Aug;103:74-91. [doi: 10.1016/j.eswa.2018.02.043]

38.   Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]

39.   Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of rmojis. PLoS One 2015;10(12):e0144296 [FREE Full text] [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]

40.   Unicode. URL: https://home.unicode.org/

41.   Emojipedia. URL: https://emojipedia.org/

42.   Gallo K. Understanding children's food-related emotions using words and emojis in the United States and Ghana. Kansas State University. 2016. URL: https://krex.k-state.edu/dspace/bitstream/handle/2097/34558/KatherineGallo2016.pdf?sequence=1

43.   Yang Y. Are you emoji savvy? Exploring nonverbal communication through emojis. Commun Teacher 2019 Mar 28;34(1):2-7. [doi: 10.1080/17404622.2019.1593472]

44. Salkind N. Krippendorff's alpha. In: Encyclopedia of Research Design. Newbury Park, California: SAGE Publications, Inc; 2010.

45. Jin J, Bai P, He W, Wu F, Liu X, Han D, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health 2020;8:152. [doi: 10.3389/fpubh.2020.00152] [Medline: 32411652]

46. Messacar D, Morissette R, Deng Z. The economic lockdown to stop the spread of COVID-19 has led to steep declines in employment and hours worked for many Canadians. Statistics Canada. 2020. URL: https://www150.statcan.gc.ca/n1/pub/45-28-0001/2020001/article/00029-eng.htm [accessed 2020-08-31]

47. Wareham J. Why some LGBT people feel uneasy at the sight of NHS rainbow flags. Forbes. 2020 May 06. URL: https://www.forbes.com/sites/jamiewareham/2020/05/06/should-the-lgbt-community-call-out-nhs-appropriation-of-rainbow-flag/ [accessed 2020-08-31]

48. Sehl K. Top Twitter demographics that matter to social media marketers. Hootsuite. 2020 May 28. URL: https://blog.hootsuite.com/twitter-demographics/ [accessed 2020-08-31]

49. Smith TW. Cross-national differences in attitudes towards homosexuality. UCLA: The Williams Institute 2011:1.

50. Martucci J. Negotiating exclusion: MSM, identity, and blood policy in the age of AIDS. Soc Stud Sci 2010 Apr;40(2):215-241. [doi: 10.1177/0306312709346579] [Medline: 20527321]

51. Capraro V, Barcelo H. The effect of messaging and gender on intentions to wear a face covering to slow down COVID-19 transmission. PsyArXiv. Preprint posted online May 11, 2020 . [doi: 10.31234/osf.io/tg7vz]

52. Operation 50/50: women's perspectives save lives. Women in Global Health. 2020. URL: https://www.womeningh.org/operation-50-50

53. Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. Lancet 2020 Mar;395(10227):846-848. [doi: 10.1016/s0140-6736(20)30526-2]

54. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]

55. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Abbreviations

API: application programming interface
F2F: face-to-face

<u>Original Paper</u>

Leveraging Internet Search Data to Improve the Prediction and Prevention of Noncommunicable Diseases: Retrospective Observational Study

Chenjie Xu[1*], MSc; Zhi Cao[1,2*], MSc; Hongxi Yang[1,3], PhD; Ying Gao[4], PhD; Li Sun[5], PhD; Yabing Hou[1], MSc; Xinxi Cao[1], MSc; Peng Jia[6,7], PhD; Yaogang Wang[1], MD, PhD

[1]School of Public Health, Tianjin Medical University, Tianjin, China

[2]Department of Epidemiology and Health Statistics, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China

[3]School of Public Health, Yale University, New Haven, CT, United States

[4]Health Management Center, Tianjin Medical University General Hospital, Tianjin, China

[5]School of Nursing, Tianjin Medical University, Tianjin, China

[6]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

[7]International Institute of Spatial Lifecourse Epidemiology, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Yaogang Wang, MD, PhD
School of Public Health
Tianjin Medical University
No. 22, Qixiangtai Road
Heping District
Tianjin, 300070
China
Phone: 86 13820046130
Email: <u>wyg@tmu.edu.cn</u>

Abstract

Background: As human society enters an era of vast and easily accessible social media, a growing number of people are exploiting the internet to search and exchange medical information. Because internet search data could reflect population interest in particular health topics, they provide a new way of understanding health concerns regarding noncommunicable diseases (NCDs) and the role they play in their prevention.

Objective: We aimed to explore the association of internet search data for NCDs with published disease incidence and mortality rates in the United States and to grasp the health concerns toward NCDs.

Methods: We tracked NCDs by examining the correlations among the incidence rates, mortality rates, and internet searches in the United States from 2004 to 2017, and we established forecast models based on the relationship between the disease rates and internet searches.

Results: Incidence and mortality rates of 29 diseases in the United States were statistically significantly correlated with the relative search volumes (RSVs) of their search terms ($P<.05$). From the perspective of the goodness of fit of the multiple regression prediction models, the results were closest to 1 for diabetes mellitus, stroke, atrial fibrillation and flutter, Hodgkin lymphoma, and testicular cancer; the coefficients of determination of their linear regression models for predicting incidence were 80%, 88%, 96%, 80%, and 78%, respectively. Meanwhile, the coefficient of determination of their linear regression models for predicting mortality was 82%, 62%, 94%, 78%, and 62%, respectively.

Conclusions: An advanced understanding of search behaviors could augment traditional epidemiologic surveillance and could be used as a reference to aid in disease prediction and prevention.

KEYWORDS

XSL•FO
RenderX

Introduction

One of the Sustainable Development Goals (SDGs) set by the United Nations General Assembly in 2015 was to reduce premature mortality from noncommunicable diseases (NCDs) by one-third by 2030 [1]. According to the World Health Statistics 2019 [2], NCDs have collectively caused 41 million deaths worldwide, equivalent to 71% of all global deaths. The majority of those deaths were caused by the following NCDs: cardiovascular disease (CVD), cancer, and diabetes. According to the statistics of the Global Burden of Disease (GBD) in 2017 [3], diabetes was the most common NCD in the United States and ischemic heart disease ranked second. In 2020, 1,806,590 new cancer cases and 606,520 cancer deaths are projected to occur in the United States [4].

In the era of social media, there is a current trend in the population for individuals to search the internet for information before they consult with specialists for recommendations [5-7]. Various social media and online communities that enable connectivity have unprecedented influence. They are expanding their reach into the health care domain [8-12]. Researchers have shown that patients with cancer, diabetes, and other chronic conditions search online before and after diagnosis [13]. According to a report by the American Community Survey [14], the percentage of households with a computer has increased almost tenfold since 1984. Like computer use, the percentage of households using the internet has also increased over time [14]. Google accounts for the vast majority of the US search engine market, reaching more than 80% of the population [15].

The application of search engine data to the field of disease surveillance is a classic case of big data application. When GFT was first launched [16], it attracted widespread attention and was followed by many scholars. Previous studies have mainly utilized Google to study outbreaks of influenza epidemics [17-19]. Other studies have tried to mine social network data (such as Facebook and Twitter) through text mining to identify patients' concerns about drugs, examine their feelings, and understand public opinion [20]. Due to the lack of real-time monitoring data of NCDs, few studies had used internet data to predict the trends of NCDs [21,22]. Our study is the first to explore a correlation among these many different types of online NCD searches with disease prevalence in the United States. We hypothesized that internet search behaviors could reflect people's awareness of NCDs, and thus the disease characteristics of NCDs (such as incidence and mortality rates) would correlate strongly with the internet search frequency. It also provides a new information channel for grasping public health concerns and promoting NCD prevention.

Methods

Disease Data

We obtained the national incidence and mortality rates for NCDs in the United States from the GBD database for 14 years (from 2004 to 2017) [3]. In this study, we initially selected 31 types of NCDs, namely diabetes mellitus, ischemic heart disease, stroke, atrial fibrillation and flutter, prostate cancer, breast cancer, lung cancer, colon and rectal cancer, malignant skin melanoma, non-Hodgkin lymphoma, uterine cancer, cardiomyopathy and myocarditis, kidney cancer, pancreatic cancer, bladder cancer, leukemia, liver cancer, stomach cancer, lip and oral cavity cancer, brain and nervous system cancer, thyroid cancer, multiple myeloma, ovarian cancer, cervical cancer, esophageal cancer, larynx cancer, gallbladder and biliary tract cancer, Hodgkin lymphoma, testicular cancer, mesothelioma, and hypertensive heart disease.

Internet Search Data

Because internet search data are updated in real time, this study mainly considered monthly internet searches from the Google Trends website from 2004 to 2018 at the national level [23]. We downloaded monthly relative search volumes for each search term of each disease. The search data were downloaded from Google Trends in December 2019.

Selection of Search Terms

Different kinds of NCDs have different search terms and each disease has a core search term. We determined the core search term based on the disease name in the GBD database. Google Trends has the function of "related searches" [24]: after entering the core search term, other search terms related to this term in the "related searches" section can be seen at the bottom of the page, which contains up to 49 related search terms and sentences. This method determines the approximate scope of a search term selection based on the object to be studied, which can avoid the subjectivity of search term selection in the research to a certain extent and can minimize the omission of core terms. After the primary selection of terms, the next step was the filtering of search terms. Three types of terms were generally filtered out in this study. The first type was terms with meanings that had nothing to do with the research object. After the primary selection, some terms were still irrelevant to the research object, even if a phrase included the object to be studied, likely because some words have multiple meanings. The second type of terms to filter out was those with a small search volume. Some of the included terms had zero search volume within the specified time frame. Because this study focused on time differences, it was required that search terms had a high search frequency throughout the entire period.

Specific search terms are listed in Multimedia Appendix 1. The selected terms were not searched in quotes. Each data point represents the relative search volumes (RSVs) of specific query terms on a normalized scale of 0 to 100. The RSVs were divided by the total searches of the particular geographic location and a particular time range it represents to compare the relative popularity of the query terms. For example, compared with the total search volumes, if a particular region had a higher number of specific query terms, its RSV would be closer to 100. Data of internet searches used in this study are publicly available, anonymous, and cannot be tracked back to identifiable individuals.

Statistical Analysis

We identified search terms for each NCD based on the above criteria. We performed the Pearson correlation analysis to evaluate the relationship between the known incidence and mortality rates of the NCDs and the RSVs to filter search terms

continuously. Finally, the terms that have no significant correlation with the research object were also deleted in the subsequent analysis.

We also considered dealing with the multicollinearity of the RSVs of the search terms. Multicollinearity refers to the distortion or inaccuracy of model estimation due to the high correlation between explanatory variables in a linear regression model. In this study, we calculated the correlation coefficients between the RSVs of the search terms of each disease.

Based on the above steps, we established multiple linear regression models for each disease, and the RSVs of multiple search terms were used for prediction and analysis.

The general form of the multiple linear regression model can be expressed as follows:

$$y = \beta_0 + \beta_1 x_2 + \beta_2 x_2 + ... + \beta_K x_k + \varepsilon,$$

where β_0, β_1, β_2,..., β_K is the parameter of the model and ε is the error term. The error term reflects the influence of random factors on y, which cannot be determined by the variability explained by the linear relationship between x_k and y. In this study, we established two regression models for each NCD, one based on the correlation between RSVs and incidence rate and the other based on the correlation between RSVs and the mortality rate.

Statistical analysis was conducted using IBM SPSS software (version 22.0), and Stata (version 15; StataCorp LLC). The statistical significance was set as $P<0.05$ (two-sided test).

Results

Correlation Analysis

Table 1 and Multimedia Appendix 2 display the correlation coefficients between the incidence rates and the RSVs of all of the selected search terms for the NCDs. They also display the correlation coefficients between the mortality rates and the RSVs. We found statistically significant correlations between the rates and the RSVs, especially for five diseases with high incidence rates: diabetes mellitus, stroke, atrial fibrillation and flutter, Hodgkin lymphoma, and testicular cancer. For Hodgkin lymphoma, the RSV of each search term was negatively correlated with the incidence rate. For some diseases, we did not find statistically significant correlations between the RSVs and the mortality and incidence rates; these search terms were excluded from the subsequent analysis. If the correlations between the rates and RSVs of all search terms of a disease were found to be not statistically significant, the disease would be excluded from the analysis. For example, prostate cancer and hypertensive heart disease were excluded from further analysis because the RSVs of all of their search terms did not correlate with incidence and mortality rates at the same time.

The RSVs of the specific search terms were correlated with the incidence rates for all of the NCDs, with *P* values less than .05 (Multimedia Appendix 3). In addition, the RSVs of the specific search terms were correlated with the mortality rates for all of the NCDs. Multimedia Appendix 4 displays the cross correlation analysis results among search terms.

Table 1. Correlation coefficients between the incidence and mortality rates of diabetes mellitus, stroke, atrial fibrillation and flutter, Hodgkin lymphoma, and testicular cancer and their relative search volumes.

	Incidence rate		Mortality rate	
	$R_{incidence}$	P value	$R_{mortality}$	P value
Search terms for diabetes mellitus				
What is diabetes mellitus type 2	0.648	<.001	–0.636	<.001
What is type 2 diabetes	0.746	<.001	–0.789	<.001
Causes of diabetes mellitus	–0.295	<.001	0.250	.001
Signs of diabetes	0.866	<.001	–0.900	.001
What is type 1 diabetes	0.729	<.001	–0.756	<.001
Hypoglycemia	–0.522	<.001	0.403	<.001
Hyperlipidemia	0.686	<.001	–0.757	<.001
Search terms for stroke				
Signs of stroke in women	0.893	<.001	–0.495	<.001
Symptoms of stroke in women	0.609	<.001	–0.663	<.001
Signs of a stroke in women	0.906	<.001	–0.447	<.001
Stroke symptoms in men	0.805	<.001	–0.655	<.001
Minor stroke	0.245	.001	0.164	.03
Symptoms of mini stroke	0.558	<.001	–0.223	.004
Signs of stroke in men	0.887	<.001	–0.392	<.001
Signs of mini stroke	0.733	<.001	–0.277	.003
What are the signs of a stroke	0.745	<.001	–0.381	<.001
Search terms for atrial fibrillation and flutter				
Atrial fibrillation	0.163	.03	0.200	.009
Afib[a]	0.953	<.001	0.936	<.001
Atrial fibrillation with rvr[b]	0.218	.004	0.209	.007
Atrial fibrillation and stroke	–0.350	<.001	–0.359	<.001
Ablation of atrial fibrillation	–0.224	.003	–0.233	.002
Heart flutter	0.880	<.001	0.878	<.001
Atrial fibrillation vs flutter	0.363	<.001	0.376	<.001
Signs of atrial fibrillation	0.202	.009	0.205	.008
Atrial flutter vs atrial fibrillation	0.230	.003	0.230	.002
Atrial flutter	0.389	<.001	0.378	<.001
Atrial flutter ecg[c]	0.179	.02	0.161	.04
Atrial flutter ekg[d]	–0.163	.03	–0.165	.03
Atrial flutter vs fibrillation	0.391	<.001	0.403	<.001
Atrial fibrillation ecg	0.435	<.001	0.430	<.001
What is atrial fibrillation	0.685	<.001	0.681	<.001
a fib[e]	0.965	<.001	0.957	<.001
Treatment of atrial flutter	–0.187	.015	–0.185	.02
What causes atrial flutter	0.328	<.001	0.331	<.001
Search terms for Hodgkin lymphoma				
Hodgkin lymphoma	–0.518	<.001	–0.404	<.001

	Incidence rate		Mortality rate	
	$R_{incidence}$	P value	$R_{mortality}$	P value
Hodgkin lymphoma cancer	–0.407	<.001	–0.316	<.001
Hodgkin lymphoma symptoms	–0.496	<.001	–0.413	<.001
Lymphoma symptoms	–0.365	<.001	–0.524	<.001
What is lymphoma	–0.773	<.001	–0.688	<.001
What is Hodgkin lymphoma	–0.635	<.001	–0.561	<.001
Hodgkins	0.823	<.001	0.801	<.001
Symptoms of Hodgkin lymphoma	–0.416	<.001	–0.346	<.001
Symptoms of lymphoma	–0.534	<.001	–0.590	<.001
Hodgkin lymphoma survival rate	–0.425	<.001	–0.357	<.001
Hodgkin lymphoma vs non Hodgkin lymphoma	–0.513	<.001	–0.432	<.001
Hodgkin vs non Hodgkin	–0.547	<.001	–0.452	<.001
Hodgkin lymphoma prognosis	–0.227	.003	–0.201	.009
B cell lymphoma	–0.303	<.001	–0.233	.002
B cell non odgkin lymphoma	–0.538	<.001	–0.493	<.001
Hodgkin lymphoma causes	–0.433	<.001	–0.363	<.001
Stage 4 lymphoma	–0.416	<.001	–0.366	<.001
Classical Hodgkin lymphoma	–0.495	<.001	–0.490	<.001
Hodgkin lymphoma stage 4	–0.415	<.001	–0.344	<.001
Search terms for testicular cancer				
Testicular cancer	–0.821	<.001	0.742	<.001
Causes of testicular cancer	–0.292	<.001	0.280	<.001
Symptoms for testicular cancer	–0.196	.01	0.248	.001
Testicular cancer ribbon	0.342	<.001	–0.275	<.001
Testicular cancer prognosis	–0.366	<.001	0.368	<.001
Testicular cancer risk factors	–0.185	.02	0.217	.005
What does testicular cancer look like	0.317	<.001	–0.254	.001
How do you get testicular cancer	0.320	<.001	–0.214	.005
Does testicular cancer spread	0.297	<.001	–0.217	.005
What are signs of testicular cancer	0.557	<.001	–0.476	<.001

[a]afib: atrial fibrillation.

[b]rvr: rapid ventricular response.

[c]ecg: electrocardiogram.

[d]ekg: electrocardiogram.

[e]a fib: atrial fibrillation.

Trends in Internet Searches, Incidence Rates, and Mortality Rates

Figure 1 shows a time series of the RSVs, incidence rates, and mortality rates for atrial fibrillation and flutter from 2004 to 2018. Trends of other diseases are displayed in Multimedia Appendix 5. Based on the correlation analysis, we predicted the incidence and mortality rates in 2018. As can be seen in the figure, the incidence rates of most diseases and RSVs fit well. A similar pattern was observed between RSVs and mortality rates. The predicted incidence and mortality rates varied with the fluctuations in the RSVs for most NCDs.

Figure 1. Trends of atrial fibrillation and flutter from 2004 to 2018. (A) Trends of incidence rate and relative search volume (RSV) of atrial fibrillation and flutter from 2004 to 2018. (B) Trends of mortality rate and RSV of atrial fibrillation and flutter from 2004 to 2018. The blue line represents the RSV of each noncommunicable disease from 2004 to 2018, the green line represents the incidence rates, the purple line represents the mortality rates, and the dotted line represents the forecast for morbidity and mortality in 2018.
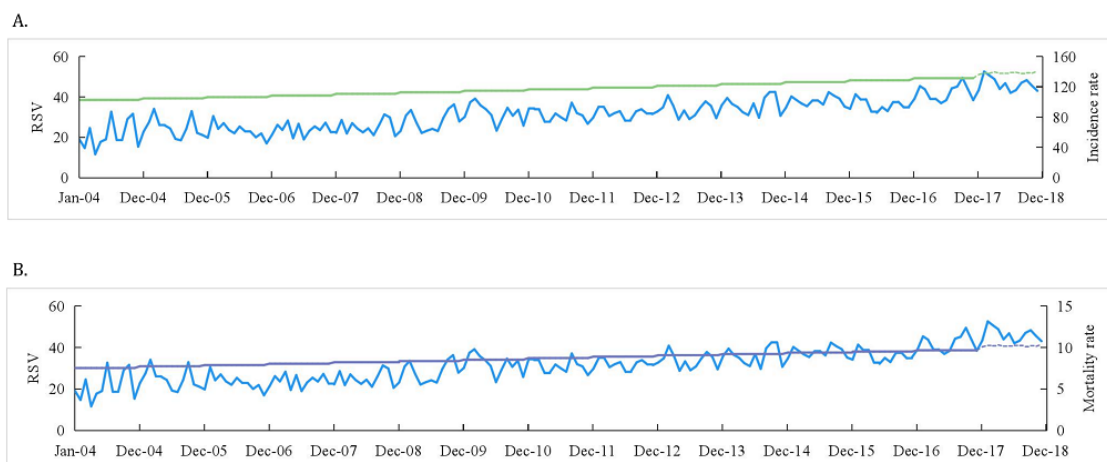
Multiple Linear Regression Models

Based on the correlations, two prediction models were established for each disease, namely the incidence prediction model and the mortality prediction model. Figure 2 shows the relationship between the independent variable (RSV for each search term) and the dependent variable (incidence and mortality rates of atrial fibrillation and flutter) in the model. The relationship between the independent variable and the dependent variable of all of the NCDs can be seen in Multimedia Appendix 6. The prediction models are shown in Multimedia Appendix 7. Table 2 shows the degree of fit of multiple linear regression models. The results of all of the NCDs are displayed in Multimedia Appendix 8. For diabetes mellitus, stroke, atrial fibrillation and flutter, Hodgkin lymphoma, and testicular cancer, the coefficient of determination of the linear regression models for predicting incidence was 80%, 88%, 96%, 80%, and 78%, respectively.

Meanwhile, the coefficient of determination of the linear regression models for predicting mortality was 82%, 62%, 94%, 78%, and 62%, respectively. From the perspective of the goodness of fit of the multiple regression prediction models of other NCDs, most of the results were close to 1.

Figure 2. (A) Scatter plot of relative search volumes and incidence rate of atrial fibrillation and flutter in the United States. (B) Scatter plot of relative search volumes and mortality rate of atrial fibrillation and flutter in the United States.

A.
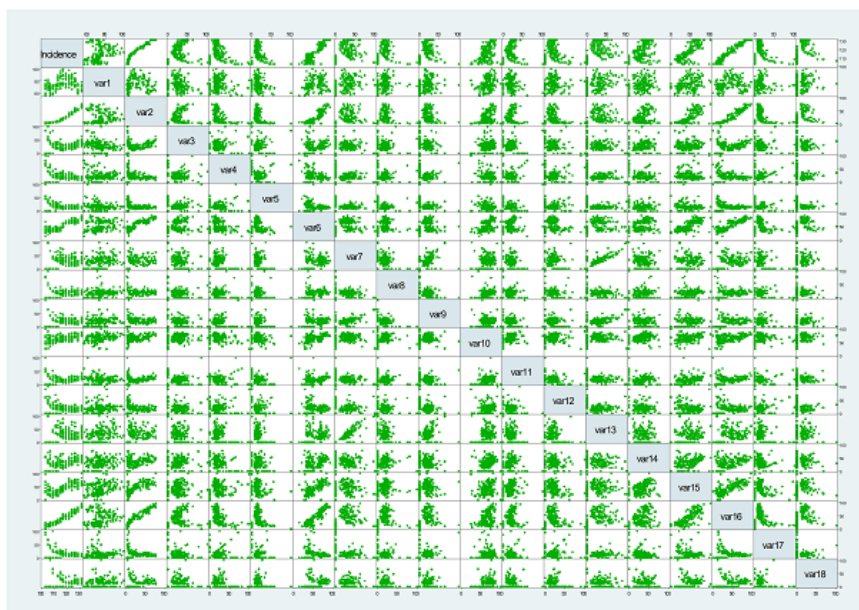
B.
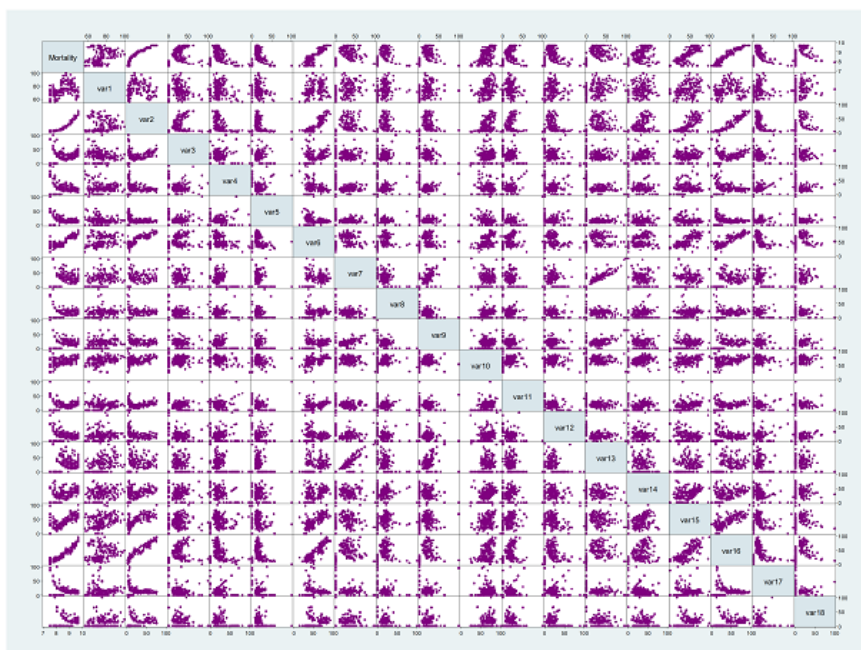
Table 2. Evaluation results of prediction models.

	Incidence rate			Mortality rate		
	R^2	Adjusted R^2	RMSE[a]	R^2	Adjusted R^2	RMSE
Diabetes mellitus	80%	80%	20.54	82%	81%	0.59
Stroke	88%	87%	1.43	62%	60%	1.15
Atrial fibrillation and flutter	96%	95%	1.91	94%	94%	0.17
Hodgkin lymphoma	80%	77%	0.10	78%	75%	0.01
Testicular cancer	78%	77%	0.01	62%	60%	0.00

[a]Root-mean-square error.

Discussion

Principal Findings

In recent years, using internet search data to detect influenza has been a research hotspot. Most of the studies utilized data sources from Google Trends or Google Flu Trends [25-27]. Although the existing literature has conducted empirical research on the correlation between search data and influenza, there is generally a lack of systematic preprocessing methods for NCDs. This study mainly focused on three types of diseases: diabetes, cancer, and CVDs. We found that the frequency of searches correlated strongly with previously reported disease epidemiology.

The choice of search terms was a key feature of this study. Based on the disease names used by the GBD, we used the "related searches" function of the Google search engine to supplement the search term and thus obtain a more comprehensive primary selection at a low cost. In terms of the search term selection, compared with Pearson correlation analysis conducted in the previous research, the use of cross-correlation analysis can also determine the time relationship between search terms and disease occurrence and determine the leading search terms in time so as to build a predictive model. The above preprocessing methods provided a better data foundation for the establishment of early warning models.

This study demonstrated the feasibility of combining historical information with search information for the early warning of NCDs, laying the foundation for future model optimization. As the most recent data publicly available from the Centers for Disease Control and Prevention is at least 2 years old while Google Trends data are available nearly instantaneously, this resource could potentially provide a more timely and cost-effective data source for public health researchers [28]. Thus, real-time internet searches could be particularly useful. Studies have shown that tracking and monitoring search behaviors, as well as text mining on social media, can provide new ways to study public concerns about NCDs and information-seeking behaviors [29-32]. Researchers have realized that people can search, understand, and evaluate health information on internet resources and that they harness the information they receive to address health problems [33]. At present, it is necessary to enhance the efficiency of prevention and auxiliary diagnosis for patients with NCDs or the general population by online information transmission. The collection of real-time relevant search data from search engines provides a new way to prevention and control NCDs.

Most types of NCDs that we examined showed statistically significant correlations, although prostate cancer and hypertensive heart disease did not. This pattern can be attributable to different reasons. First, NCDs are highly prevalent. As people are paying more attention to these diseases, the self-management consciousness of patients with certain NCDs has been improving. Second, the monthly rates of diagnosis of NCDs are missing. Third, this pattern could be partly influenced by active public health campaigns, which may broadly increase search volume regardless of disease metrics. Between 2007 and 2014, the incidence of cancer in men in the United States declined rapidly, and then remained stable until 2016, which was related to changes in screening strategies for colon, rectal, and prostate cancer [4]. Doctors and scientists in the United States have noticed that previous prostate-specific antigen (PSA)-based screenings may lead to an overdiagnosis for prostate cancer and have reduced the use of PSA for prostate cancer screening. A previous study showed that fewer older adults in the United States are having heart attacks, and among those who do have them, more are surviving from them [34]. The country has made continuous efforts to prevent heart attacks and improve patient care. Health insurance agencies, the American Heart Association, and other research organizations—as well as a large number of researchers, clinicians, and public health experts—are committed to reducing risk by promoting a healthy lifestyle. At the same time, the recommendations for the secondary prevention of heart disease are more common and standardized. The extensive development of angiography and other related technologies, and the deepening of people's understanding of CVDs, will affect the frequency of retrieval of related diseases on the internet. This trend can also be seen in our prediction model. The incidence of lung cancer in the United States has also continued to decline due to the continuous implementation of antismoking activities and effective control of the smoking rate. Fourth, we extracted internet search data from as early as 2004. In the past, individuals with the highest risk of NCDs often had limited access to the internet. To date, the internet has ushered in great changes. The number of internet users is currently growing at a rate of more than 11 new users per second, bringing the total number of new users per day to an astonishing 1 million. More and more people tend to search for health-related information online. Our results also indicated that the coefficient of determination and adjusted coefficient of determination of the

regression models for diabetes, stroke, atrial fibrillation and flutter, Hodgkin lymphoma, and testicular cancer were higher than for the other diseases studied, indicating that the regression models are better fit to the incidence and mortality rates of these NCDs. This might mean that online search behaviors and volumes can help health professionals to conduct near real-time monitoring of NCDs.

Although the method of using internet search data to predict influenza has made great progress in real time, it still lacks in accuracy. Since 2011, GFT has been overestimating the number of influenza-like illnesses (ILIs), especially the forecast of the peak season of influenza. In 2013, the forecast deviation was even as high as 140%, which may indicate that there is a certain gap in the forecast based on internet data alone. In the traditional methods of influenza prediction, although the manually collected ILI monitoring samples are lagging behind, they are more accurate because of rigorous scientific experiments. Therefore, internet data cannot completely replace the traditional data collection methods but should be used as a complement to the latter.

Limitations

Our study has some unavoidable limitations. First, because the data from internet searches are public and anonymous, we could not determine who conducted search activities in this study, and the limited number of search terms could not fully represent the search preferences of all people. Second, the use of Google Trends cannot be fully representative of the overall population, since only individuals with access to the internet can be accounted for. Third, we could only obtain the annual incidence and mortality data for NCDs, but not the data with smaller granularity, which might have affected the accuracy of the prediction model. Fourth, as the search algorithm of Google Trends is dynamic, we could not retrieve the original RSVs, and the RSVs of the same search term obtained in different time periods were different. Fifth, the spatiotemporal data of search engines have many limitations, such as high noise and uncertainty. We hope to find ways to identify and reduce bias in search engine data before we utilize web-based data to provide useful information.

Public Health Implications

With the widespread use of internet searches, our study found a correlation between the RSVs and the incidence and mortality rates of the NCDs. This indicates that the search engine data can be used in the early warning and prevention of NCDs, such as diabetes, cancer, and CVDs. We should make good use of such data, especially when the traditional registry data are insufficient or unavailable.

Authors' Contributions

YW directed the study. YH downloaded all the original Google Trends data in the United States. CX and ZC processed and analyzed the data, and then developed the first manuscript draft. All authors critically revised the manuscript, and all authors critically reviewed, contributed to, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms of all the noncommunicable diseases.
[XLSX File (Microsoft Excel File), 35 KB - jmir_v22i11e18998_app1.xlsx]

Multimedia Appendix 2
Correlation analysis among the relative search volumes, incidence rates, and mortality rates.
[XLSX File (Microsoft Excel File), 34 KB - jmir_v22i11e18998_app2.xlsx]

Multimedia Appendix 3
Correlation coefficients among internet searches, incidence rates, and mortality rates in the United States (contains search terms that were ultimately included in the study).
[DOC File , 584 KB - jmir_v22i11e18998_app3.doc]

Multimedia Appendix 4
Cross-correlation analysis results among search terms.
[XLSX File (Microsoft Excel File), 38 KB - jmir_v22i11e18998_app4.xlsx]

Multimedia Appendix 5

XSL•FO
RenderX

Trends of all the noncommunicable diseases.
[DOC File , 1062 KB - jmir_v22i11e18998_app5.doc]

Multimedia Appendix 6
Relationship between the independent variable and the dependent variable of all the noncommunicable diseases.
[PDF File (Adobe PDF File), 30670 KB - jmir_v22i11e18998_app6.pdf]

Multimedia Appendix 7
Prediction models.
[DOC File , 40 KB - jmir_v22i11e18998_app7.doc]

Multimedia Appendix 8
Evaluation results of prediction models for all the chronic diseases.
[DOC File , 81 KB - jmir_v22i11e18998_app8.doc]

References

1. Nugent R, Bertram MY, Jan S, Niessen LW, Sassi F, Jamison DT, et al. Investing in non-communicable disease prevention and management to advance the Sustainable Development Goals. The Lancet 2018 May;391(10134):2029-2035. [doi: 10.1016/s0140-6736(18)30667-6]
2. WHO. World health statistics 2019. URL: https://www.who.int/gho/publications/world_health_statistics/en/ [accessed 2020-01-17]
3. Global Burden of Disease. GBD results tool. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2020-01-17]
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA A Cancer J Clin 2020 Jan 08;70(1):7-30. [doi: 10.3322/caac.21590]
5. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-Based Interventions to Improve Mental Health, General Caregiving Outcomes, and General Health for Informal Caregivers of Adults With Chronic Conditions Living in the Community: Rapid Evidence Review. J Med Internet Res 2017 Jul 28;19(7):e263. [doi: 10.2196/jmir.7564]
6. Weber W, Reinhardt A, Rossmann C. Lifestyle Segmentation to Explain the Online Health Information–Seeking Behavior of Older Adults: Representative Telephone Survey. J Med Internet Res 2020 Jun 12;22(6):e15099. [doi: 10.2196/15099]
7. Gedefaw A, Yilma TM, Endehabtu BF. Information Seeking Behavior About Cancer and Associated Factors Among University Students, Ethiopia: A Cross-Sectional Study. Cancer Manag Res 2020;12:4829-4839 [FREE Full text] [doi: 10.2147/CMAR.S259849] [Medline: 32606972]
8. Alioshkin Cheneguin A, Salvat Salvat I, Romay Barrero H, Torres Lacomba M. How good is online information on fibromyalgia? An analysis of quality and readability of websites on fibromyalgia in Spanish. BMJ Open 2020 Jul 05;10(7):e037065. [doi: 10.1136/bmjopen-2020-037065]
9. Heynsbergh N, Botti M, Heckel L, Livingston PM. Caring for the person with cancer: Information and support needs and the role of technology. Psycho-Oncology 2018 Apr 20;27(6):1650-1655. [doi: 10.1002/pon.4722]
10. Tonsaker T, Bartlett G, Trpkov C. Health information on the Internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
11. Teh J, Op't Hoog S, Nzenza T, Duncan C, Wang J, Radojcic M, et al. Penile cancer information on the internet: a needle in a haystack. BJU Int 2018 Oct 29;122:22-26. [doi: 10.1111/bju.14532]
12. Printz C. Patients with cancer seeking treatment information on the internet face challenges. Cancer 2017 May 19;123(11):1886-1887. [doi: 10.1002/cncr.30773]
13. Heynsbergh N, Botti M, Heckel L, Livingston PM. Caring for the person with cancer: Information and support needs and the role of technology. Psycho-Oncology 2018 Apr 20;27(6):1650-1655. [doi: 10.1002/pon.4722]
14. Camille R. Computer and Internet use in the United States: 2016. URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2020-09-17]
15. StatCounter. Search engine market share. URL: https://gs.statcounter.com/search-engine-market-share [accessed 2020-01-01]
16. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634]
17. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. Nat Commun 2019 Jan 11;10(1):147. [doi: 10.1038/s41467-018-08082-0]
18. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. The Lancet Infectious Diseases 2014 Feb;14(2):160-168. [doi: 10.1016/s1473-3099(13)70244-5]
19. Zhang Y, Bambrick H, Mengersen K, Tong S, Hu W. Using Google Trends and ambient temperature to predict seasonal influenza outbreaks. Environment International 2018 Aug;117:284-291. [doi: 10.1016/j.envint.2018.05.016]

20. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns. JAMA Ophthalmol 2016 Sep 01;134(9):1024. [doi: 10.1001/jamaophthalmol.2016.2267]

21. Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of Search Engine Queries for Chest Pain With Coronary Heart Disease Epidemiology. JAMA Cardiol 2018 Dec 01;3(12):1218-1221. [doi: 10.1001/jamacardio.2018.3459]

22. Toosi B, Kalia S. Seasonal and Geographic Patterns in Tanning Using Real-Time Data From Google Trends. JAMA Dermatol 2016 Feb 01;152(2):215-217. [doi: 10.1001/jamadermatol.2015.3008]

23. Google. Google Trends. URL: https://trends.google.com/trends/ [accessed 2019-10-01]

24. Trends Help. Google Trends. URL: https://support.google.com/trends#topic=6248052 [accessed 2020-11-08]

25. Kandula S, Shaman J. Reappraising the utility of Google Flu Trends. PLoS Comput Biol 2019 Aug 2;15(8):e1007258. [doi: 10.1371/journal.pcbi.1007258]

26. Pollett S, Boscardin WJ, Azziz-Baumgartner E, Tinoco YO, Soto G, Romero C, et al. Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection. Clin Infect Dis 2016 Sep 26;64(1):34-41. [doi: 10.1093/cid/ciw657]

27. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci USA 2015 Nov 09;112(47):14473-14478. [doi: 10.1073/pnas.1515373112]

28. Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, et al. Association Between Cancer Incidence and Mortality in Web-Based Data in China: Infodemiology Study. J Med Internet Res 2019 Jan 29;21(1):e10677. [doi: 10.2196/10677]

29. Deiner MS, McLeod SD, Wong J, Chodosh J, Lietman TM, Porco TC. Google Searches and Detection of Conjunctivitis Epidemics Worldwide. Ophthalmology 2019 Sep;126(9):1219-1229. [doi: 10.1016/j.ophtha.2019.04.008]

30. Phillips CA, Barz Leahy A, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. J Med Internet Res 2018 Jan 08;20(1):e6. [doi: 10.2196/jmir.8870]

31. Huang X, Baade P, Youlden DR, Youl PH, Hu W, Kimlin MG. Google as a cancer control tool in Queensland. BMC Cancer 2017 Dec 4;17(1):816. [doi: 10.1186/s12885-017-3828-x]

32. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and Understanding Communities Using Twitter to Connect About Depression: Cross-Sectional Study. JMIR Ment Health 2018 Nov 05;5(4):e61. [doi: 10.2196/mental.9533]

33. Norman C. eHealth Literacy 2.0: Problems and Opportunities With an Evolving Concept. J Med Internet Res 2011 Dec 23;13(4):e125. [doi: 10.2196/jmir.2035]

34. Krumholz HM, Normand ST, Wang Y. Twenty-Year Trends in Outcomes for Older Adults With Acute Myocardial Infarction in the United States. JAMA Netw Open 2019 Mar 15;2(3):e191938. [doi: 10.1001/jamanetworkopen.2019.1938]

Abbreviations

CVD: cardiovascular disease
GBD: Global Burden of Disease
GFT: Google Flu Trends
ILIs: influenza-like illnesses
NCDs: noncommunicable diseases
PSA: prostate-specific antigen
RSV: relative search volume
SDGs: Sustainable Development Goals

XSL•FO

RenderX

Original Paper

The COVID-19 Infodemic: Infodemiology Study Analyzing Stigmatizing Search Terms

Zhiwen Hu[1*], PhD; Zhongliang Yang[2*], PhD; Qi Li[2*], BA; An Zhang[3], PhD

[1]School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

[2]Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing, China

[3]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhiwen Hu, PhD
School of Computer and Information Engineering, Zhejiang Gongshang University
No 18 Xuezheng Street
Higher Education Zone
Hangzhou, 310018
China
Phone: 86 0571 28008316
Email: huzhiwen@zjgsu.edu.cn

Abstract

Background: In the context of the COVID-19 infodemic, the global profusion of monikers and hashtags for COVID-19 have found their way into daily communication and contributed to a backlash against China and the Chinese people.

Objective: This study examines public engagement in crisis communication about COVID-19 during the early epidemic stage and the practical strategy of social mobilization to mitigate the infodemic.

Methods: We retrieved the unbiased values of the top-ranked search phrases between December 30, 2019, and July 15, 2020, which normalized the anonymized, categorized, and aggregated samples from Google Search data. This study illustrates the most-searched terms, including the official COVID-19 terms, the stigmatized terms, and other controls, to measure the collective behavioral propensities to stigmatized terms and to explore the global reaction to the COVID-19 epidemic in the real world. We calculated the ratio of the cumulative number of COVID-19 cases to the regional population as the cumulative rate (R) of a specific country or territory and calculated the Gini coefficient (G) to measure the collective heterogeneity of crowd behavior.

Results: People around the world are using stigmatizing terms on Google Search, and these terms were used earlier than the official names. Many stigmatized monikers against China (eg, "Wuhan pneumonia," G=0.73; "Wuhan coronavirus," G=0.60; "China pneumonia," G=0.59; "China coronavirus," G=0.52; "Chinese coronavirus," G=0.50) had high collective heterogeneity of crowd behavior between December 30, 2019, and July 15, 2020, while the official terms "COVID-19" (G=0.44) and "SARS-CoV-2" (G=0.42) have not become de facto standard usages. Moreover, the pattern of high consistent usage was observed in 13 territories with low cumulative rates (R) between January 16 and July 15, 2020, out of 58 countries and territories that have reported confirmed cases of COVID-19. In the scientific literature, multifarious naming practices may have provoked unintended negative impacts by stigmatizing Chinese people. The World Health Organization; the United Nations Educational, Scientific and Cultural Organization; and the media initiated campaigns for fighting back against the COVID-19 infodemic with the same mission but in diverse voices.

Conclusions: Infodemiological analysis can articulate the collective propensities to stigmatized monikers across search behaviors, which may reflect the collective sentiment of backlash against China and Chinese people in the real world. The full-fledged official terms are expected to fight back against the resilience of negative perceptual bias amid the COVID-19 epidemic. Such official naming efforts against the infodemic should be met with a fair share of identification in scientific conventions and sociocultural paradigms. As an integral component of preparedness, appropriate nomenclatures should be duly assigned to the newly identified coronavirus, and social mobilization in a uniform voice is a priority for combating the next infodemic.

XSL•FO
RenderX

Introduction

Background

The COVID-19 infodemic, associated with the COVID-19 outbreak, is getting some attention by researchers and policy makers [1,2]. On December 31, 2019, the novel coronavirus (2019-nCoV) disease was first reported from Wuhan City in China. With the spread of COVID-19, a massive information epidemic has undermined and disrupted global efforts to fight against COVID-19. However, the infodemic challenge does not receive enough attention in publications to fully understand it, and its unique risks have only begun to be explored [3]. The infodemic is partly characterized by a high information supply of information of variable quality, and a demand for timely and trustworthy information about 2019-nCoV [2,4-6].

On one hand, the global profusion of running headlines often inscribe fear, prejudice, disgust, and hostility into hashtags and monikers, branding discrimination and stoking panic [7-9]. Those monikers and morbid contents always combine with each other in the epicenter of an infodemic, wherein one sheds light on the social contagion of the other. The past few months have witnessed a growth of stigmatized monikers, which have found their way into daily communication and contributed to the backlash against China, Chinese people, and Asians in general [10-13]. Even worse, scientists frequently used similar monikers in the pandemic paper tsunami and exacerbated the situation [14,15].

On the other hand, individual perceptual bias could lead to insufficient or excessive information seeking, which further results in collective perceptual biases [16]. Pithy proper names are expected to be powerful in the ongoing campaign against the infodemic. As an integral component of preparedness, appropriate nomenclatures should be duly assigned to the newly identified coronavirus, which causes respiratory tract disease in humans and has had an impact on public health. So far, there are no universally accepted names, either for academic-industrial usage or consistency with international virus taxonomy.

To address such issues, we used available metadata to unfold the nature of the COVID-19 epidemic and the infodemic in this study [17].

Study Objectives

An *inappropriate* official nomenclature might fuel the infodemics unconsciously. In recent years, humans have witnessed several outbreaks of infectious diseases caused by viruses, with common names given by stakeholders. Each round of naming practice is not always successful. As a case in point, some strongly held, but flawed, names such as "Middle Eastern Respiratory Syndrome" [18] and "Swine flu" were accused of unintentional social and negative economic impacts by stigmatizing certain industries or communities. "Swine flu," an influenza strain that is known to have originated in pigs, resulted in financial damage to farmers, despite there being no evidence

that it could be spread via pork consumption [19,20]. Since these incidents, in May 2015, the World Health Organization (WHO) released some naming conventions for the naming of new human diseases [21].

Infodemics long predate COVID-19 [22]. Unfortunately, with the spread of the COVID-19 epidemic, another massive infodemic has spread virally over the world with recurring episodes [23,24]. Previous evidence suggests that the internet, by its nature, could amplify and relay such infodemics swiftly worldwide, cause exaggerated panic, and progressively worsen stigmatization against people in the epicenter of an outbreak [23-26]. In the ongoing infodemic, Corona beer is being affected by the name's similarity to the deadly coronavirus. In fact, the Mexican brand originated back in 1925 before the first strain of coronavirus was discovered and named. To address such challenges, the WHO declared this infodemic as the "2019-nCoV infodemic" on February 2, 2020 [4].

Based on a critical review, this study aims to take samples of the trillions of Google searches in connection with COVID-19 from December 30, 2019, to July 15, 2020, to address the following research issues:

- Did public engagement in the crisis communication reflect the collective sentiment of backlash against China and Chinese people in the real world? What were the global and geographical patterns of collective behavioral propensities to stigmatized monikers and the official terms?
- Were informed scientists well versed in the naming conventions to minimize unintentional negative impacts?
- What is the cohesive strategy of social mobilization to fight back against the COVID-19 infodemic?

Methods

Infodemiology, a term coined by Eysenbach, is an emerging transdisciplinary area of research studying the epidemiology of information to address the pressing concerns of public health and policy decisions [2,17,27,28]. The transdisciplinary nature of infodemiology can be found in Multimedia Appendix 1.

As of February 29, 2020, COVID-19 has spread to 60 countries and territories. Of these, the WHO published the number of cumulative cases in 54 member states on February 29, 2020, as well as Hong Kong, Macao, and Taiwan. We retrieved the cumulative cases of three nonmember states (Iceland, Azerbaijan, and Monaco) from their official websites. The corresponding total populations of 2019 come from the United Nations [29]. The cumulative rate R can be viewed as a ratio of the cumulative number of COVID-19 cases to the regional population: $R = i/p$, where i is the number of confirmed infections in a given country or territory and p is the national or regional population.

In this study, we retrieved metadata from three information sources: an electronic books corpus (Google Books Ngram Corpus [GBNC]), journals (Web of Science [WoS] and

XSL•FO

RenderX

PubMed), and the internet (Google Trends Index [GTI]), and used them to facilitate subsequent analysis.

First, the GBNC provides a unique linguistic landscape that benefits from centuries of rich grammatical and lexical resources as well as cultural contexts [30]. Each taxonomic procedure often begins with a search through tomes for comparative morphologic variations to crystallize a pithy and appropriate neologism. For example, the earliest usage of *coronavirus* and *coronaviruses* could provide an insightful and compelling argument for the historical story and help us understand the essence of the name. The diachronic discourse of *coronavirus*, *coronaviruses*, *Coronaviridae,* and *Nidovirales* promises to articulate the unfolding chronological historical time scale from when these terms were first used [31].

Second, WoS and PubMed are publication databases with rich structural metadata. Scientometric analysis also promises to articulate the unfolding chronological picture of infodemiology. Under the umbrella of infodemiological scenarios, coupled with GBNC, scientometric analysis on diachronic discourses of pertinent keywords and phrases could reflect the historical milestones and the status quo in the field of human coronaviruses (HCoVs) research.

Third, the GTI part of this study commenced on December 30, 2019, and involved daily data collection worldwide until July 15, 2020, in 60 countries and territories that have reported confirmed cases of COVID-19 as of February 29, 2020. In the ongoing COVID-19 infodemic, stigmatized monikers against controls are ideal indicators of negative bias, and the unbiased and normalized GTIs were employed to determine their populational usages across various regions over time to characterize collective perceptual biases. GTI provides knowledge dissemination metrics for query incidences of relevant keywords and phrases, since about 63% of users on the internet use Google to search for ubiquitous information [32]. Three prominent strengths of GTI reflect the global reactions to major events in the real world: (1) vast data sets that include trillions of random searches; (2) unbiased sampling from the anonymized, categorized, and aggregated raw data; and (3) normalized indexes reconciling the dynamic nature of search volumes and the different population ratios in different regions. Therefore, the dynamic spatiotemporal pattern of GTI provides a unique lens into collective behavioral propensities of crowd behavior and demographic perception of a social contagion.

Intensive information seeking or avoidance choices may reinforce people's cognition in both positive and negative ways, which is proactively rewarded by the feedback of the targeted information available [16]. In this study, a code scheme of subjective searches in daily communication was designed to fit three inclusion criteria: (1) top-ranked search interests, (2) formal and with complete spelling, (3) consistent with the global participants as much as possible. The collective behavioral propensities to the stigmatized terms were measured and compared with that of the control groups (the official terms and their counterparts). Herein, collective behavioral propensities to stigmatized terms directly represented the latent tendency to gather and interpret health care information available, which

also reflected collective perceptual biases on preconceived judgements and social contagion [16,33-35].

To demonstrate the collective behavioral propensities, descriptive analysis and formal statistical analysis were carried out. For the descriptive analysis, the daily indexes of the relative search term volumes were separately mapped on a six-color rendering scale from 0 to 100. The earliest day of each terms debut was tracked and identified. Next, we introduced the Gini coefficient to measure the collective heterogeneity of crowd behavior. We denoted the search record of term X in the Google Trends data set as $X = \{x_1, x_2,...x_n\}$, where x_i represents the (normalized) search frequency of term X at the i-th time step. The mean value of X is:

The Gini coefficient G for term X can then be calculated as follows:

When the element values in X are equal, the Gini coefficient G takes the minimum value of 0, and when $x_i=0$ for $i = 1, ..., n-1$ and $x_n=1$, G will approach the maximum value of 1. The smaller the G is, the lower the collective heterogeneity of crowd behavior is and the higher the homogeneity of individual behaviors are. Conversely, it indicates that people are divided in the consistency of individual behaviors.

Results

The Enigmatic Nature of HCoVs Puts People on Edge

It is necessary to take a glimpse into the hierarchical Linnaean category of emerging coronaviruses [36,37]. As an international authoritative body, the WHO is responsible for naming new human infectious diseases. In 1966, the International Committee on Nomenclature of Viruses (ICNV) was established with the mission of introducing some degree of order and consistency into the naming of viruses. In 1973, the ICNV became the International Committee on Virus Taxonomy (ICTV), a global authority on the designation and naming of viruses. There are seven strains of HCoVs—HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, severe acute respiratory syndrome–related coronavirus (SARS-CoV), Middle East respiratory syndrome–related coronavirus (MERS-CoV), and SARS-CoV-2—known to cause the common cold as well as more severe respiratory disease. Of those, HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 are routinely responsible for mild respiratory illnesses like the common cold but can cause severe infections in immunocompromised individuals, while the others have caused more severe diseases [38].

The diachronic discourse of *coronavirus* and *coronaviruses* in the English corpus unveils that there was a mild increase in the numbers of printed books dealing with HCoVs after the initial description of coronaviruses in 1968 [39-41]. The discovery of the novel strain had stimulated a new wave of research into coronavirus and the diseases it causes. Furthermore, meta-analysis results from WoS and PubMed indicated that the

known knowledge remains off-limits in the field of combating emerging HCoVs [42]. The WHO declared the 2019-nCoV outbreak a Public Health Emergency of International Concern (PHEIC) on January 30, 2020. This is the sixth time the WHO has declared a PHEIC since the International Health Regulations (IHR) came into force in 2005. Before COVID-19, there have been five global health emergencies since such declaration was formalized: swine flu (2009), polio (2014), Ebola (2014 and 2019), and Zika (2016). The detailed descriptions of diachronic discourse analysis and scientometric analysis in this study can be found in Multimedia Appendix 2.

SARS-CoV-2 is the seventh identified coronavirus that can cause diseases of the respiratory tract via human-to-human transmission. It caused a mysterious pneumonia outbreak that is spreading far more quickly than the SARS-CoV and MERS-CoV diseases [1,43,44], even though the epicenter of the outbreak was locked down to curb the pandemic spread [42,45]. Presently, its underlying mechanism of clinical severity is yet to be determined, although many fatal cases have occurred [46].

On the one hand, the outbreaks of SARS-CoV, HCoV-HKU1, and SARS-CoV-2 were initially linked to China and lead people into the deep-rooted impression of China as an unsanitary entity. Chinese "wet markets" have been widely depicted as unsanitary

hot spots for the transmission of zoonotic diseases [15,47-56]. Moreover, China is inevitably vulnerable to be accused of lax epidemiological control over HCoVs [57-63].

On the other hand, the enigmatic nature of HCoVs and the many unknowns about these epidemics have put people on edge. Information overload always follows closely behind the epidemics caused by HCoVs, especially in the age of the internet [64]. This enigmatic nature deepens people's anxiety in a way that makes them respond to provocative online posts, whether intentional or not.

Collective Behavioral Propensities in the Public

We further examined what people were interested in and curious about with COVID-19. Google Trends showed the most-searched interest in the official terms (COVID-19, 2019-nCoV, SARS-CoV-2, and novel coronavirus pneumonia [NCP]), the stigmatized terms (Wuhan coronavirus, China coronavirus, Chinese coronavirus, Wuhan pneumonia, and China pneumonia), and other counterparts (novel pneumonia and novel coronavirus) from December 30, 2019, to July 15, 2020 (Figures 1 and 2). Those dynamic searches are indicators of collective behaviors across various regions over time. The detailed descriptions of the code scheme of multifarious naming practices can be found in Multimedia Appendix 3.

Figure 1. Calendar illustration on the relative search interest of the COVID-19 infodemic in the context of the COVID-19 epidemic (as of 15 July 2020; part 1).
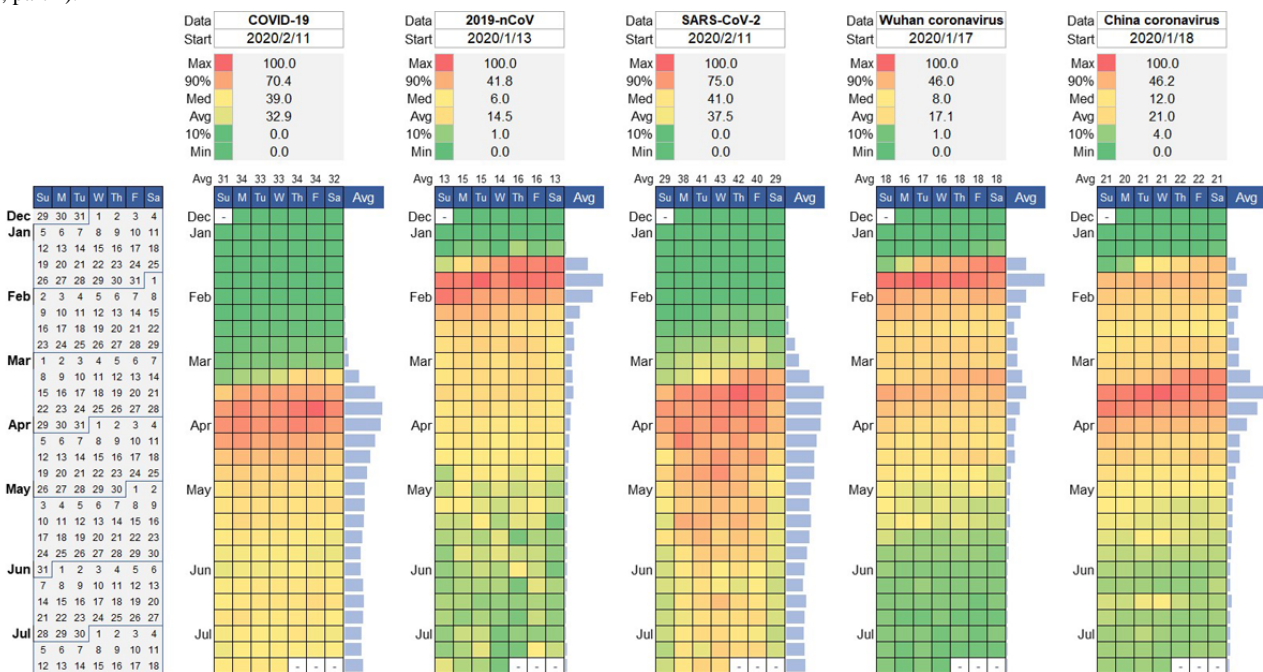
People around the world are divided in their own opinions on the internet and in daily communications. For the descriptive analysis, a striking feature was that some stigmatized monikers had comparatively high frequencies of collective usage. Being echoed by daily responses, the negative and lasting consequences pinpoint that those stigmatized names might have contributed to the recent backlash against China and Chinese people.

For the formal statistical analysis, first, the Gini coefficients of the stigmatized terms (eg, "Wuhan pneumonia," G=0.73;

"Wuhan coronavirus," G=0.60; "China pneumonia," G=0.59; "China coronavirus," G=0.52; "Chinese coronavirus," G=0.50) are significantly higher than those of the official terms (eg, "COVID-19," G=0.44; "SARS-CoV-2," G=0.42; "Novel Coronavirus Pneumonia," G=0.46) and other controls ("novel pneumonia," G=0.45 and "novel coronavirus," G=0.49). This finding strongly indicates that the homogeneity of individual propensities to stigmatized monikers are lower than the official terms and the neutral names. The vulnerable population are highly susceptible to external negative sentiments. The 2019

novel coronavirus is thought to have originated in China, this misunderstanding may have led to the high usage of "Wuhan coronavirus," "China coronavirus," "Chinese coronavirus," "Wuhan pneumonia," "China pneumonia," and other stigmatized monikers, even after July 2020. Second, after January 15, multifarious stigmatized monikers against Chinese people have prevailed in the public. "COVID-19" (*G*=0.44) took over from the premature name "2019-nCoV" (*G*=0.63), the latter finishing around February 28. Third, a notable pattern was observed after the announcements of the terms "COVID-19" and "SARS-CoV-2" (the collective usage of "SARS-CoV-2" has failed to match that of "COVID-19" in the public). The official terms "COVID-19" and "SARS-CoV-2" have not become the de facto standard usages. However, in the long run, the gradual increase in official names would be beneficial to correct ethnic stigmatization.

Figure 2. Calendar illustration on the relative search interest of the COVID-19 infodemic in the context of the COVID-19 epidemic (as of July 15, 2020; part 2).
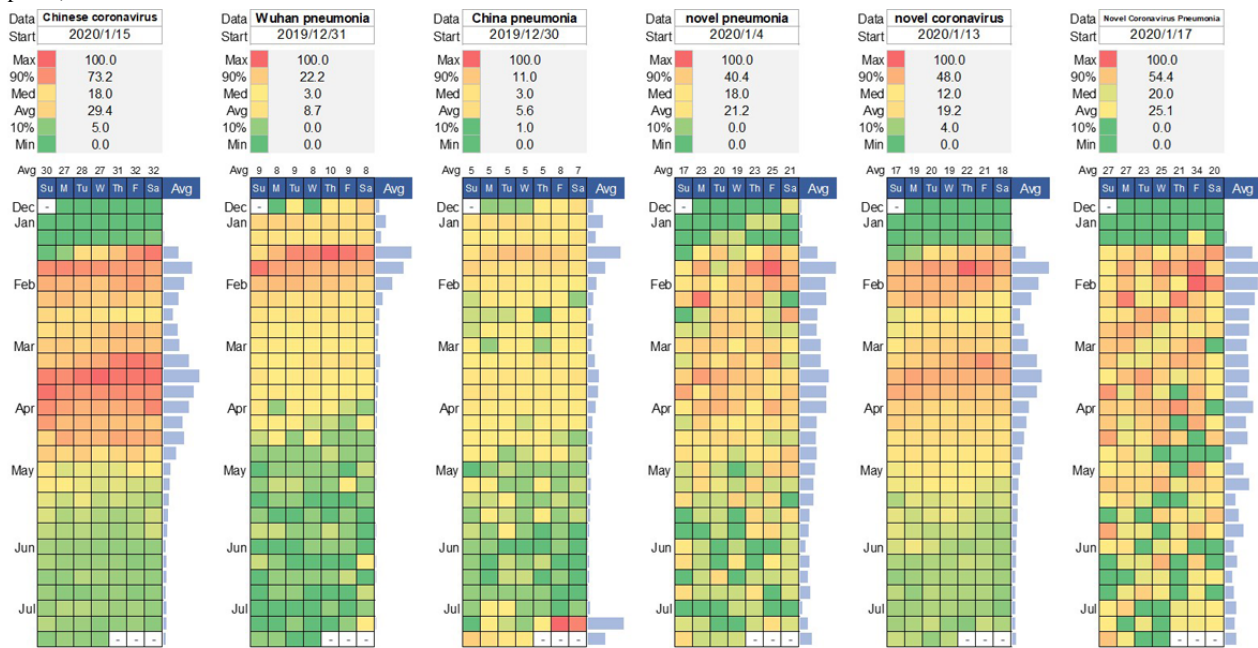
To further examine the demographical perceptions of collective behavioral propensities in the ongoing infodemic, we characterized the relationship between the geographical interest of stigmatized monikers and the cumulative rate of 58 countries and territories in which confirmed cases of COVID-19 have been reported. The results clearly unveil that people in Egypt, Greece, New Zealand, the United Kingdom, the United States, Canada, Finland, Russia, the Philippines, Denmark, Vietnam, Nepal, and Mexico prefer to use stigmatized monikers against Chinese people in comparison with other counterparts (Figure 3). As of February 29, 2020, up to 60 countries and territories have reported confirmed cases of COVID-19, including Taiwan, Iceland, Azerbaijan, and Monaco. There is no metadata available for San Marino and Monaco in Google Trends, but the geographical interest of stigmatized monikers against Chinese people in the other 58 territories was normalized by median volume to compare with each other. The cumulative rate is the ratio of the confirmed cases to the total populations in the countries or territories (Multimedia Appendix 4).

To characterize the patterns behind such collective perceptual biases, we further scrutinized the geographical interest of stigmatized monikers against China in 13 territories with low cumulative rates over time (Figure 4). This illustration unveils the geographical interest of stigmatized monikers in 13 territories (Egypt, Greece, New Zealand, the United Kingdom, the United States, Canada, Finland, Russia, the Philippines, Denmark, Vietnam, Nepal, and Mexico) between January 1,

2020, and February 29, 2020. The median volumes of the corresponding search queries showed the trend of collective behavioral propensities over time. Comparatively, some stigmatized monikers against China saw high frequencies after January 16. The substantial pattern of the high consistent curve indicates that negative perceptual bias has been observed in the perception of the natural origin of COVID-19 in the public.

As a co-occurrence perceptual phenomena, the illustration could corroborate that people have used stigmatized monikers with high frequencies in these territories after January 16, 2020. People have a negative bias in their perception of COVID-19's natural origin in these regions. Moreover, people hold negative perceptions of the authoritative responses in many countries [37,42,65,66]. The prognostic significance of our findings is that such approaches are expected to cause a psychological typhoon eye effect—a paradoxical phenomenon that the respondents closer to the epicenter of the pandemic appear to be the least concerned by the imminent risks—in the near future.

People hold negative perceptual bias in their perception of COVID-19's natural origin in 58 countries and territories with low cumulative rates (*R*; Figure 5). The substantial pattern was observed on the geographical map of 58 countries and territories with low cumulative rates and negative perceptual bias in the perception of the natural origin of COVID-19 in the public. In the sociocultural setting with a relatively complex context beyond the epidemiological dimension, this panoramic map approach allowed us to better understand the prevalence and

severity of the COVID-19 infodemic throughout the regions, comparatively. This finding reminds us that policy makers should learn from best practices to reduce deliberate infodemic risks, providing resources for knowledge and expertise in the academic sphere as well as in the public.

Figure 3. Four-quadrant diagram of the relationship between the geographical interest of stigmatized monikers and cumulative rate.
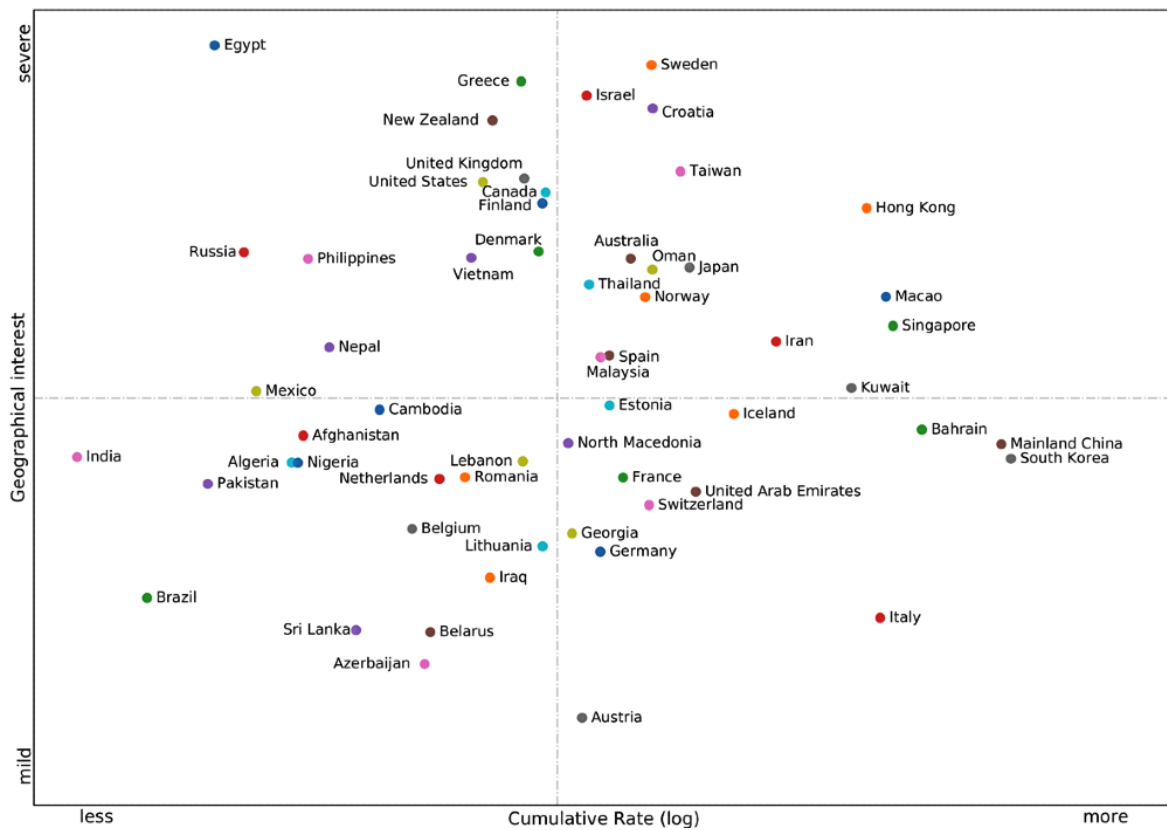
Figure 4. The dynamic interest of stigmatized monikers against China in territories with low cumulative rates.
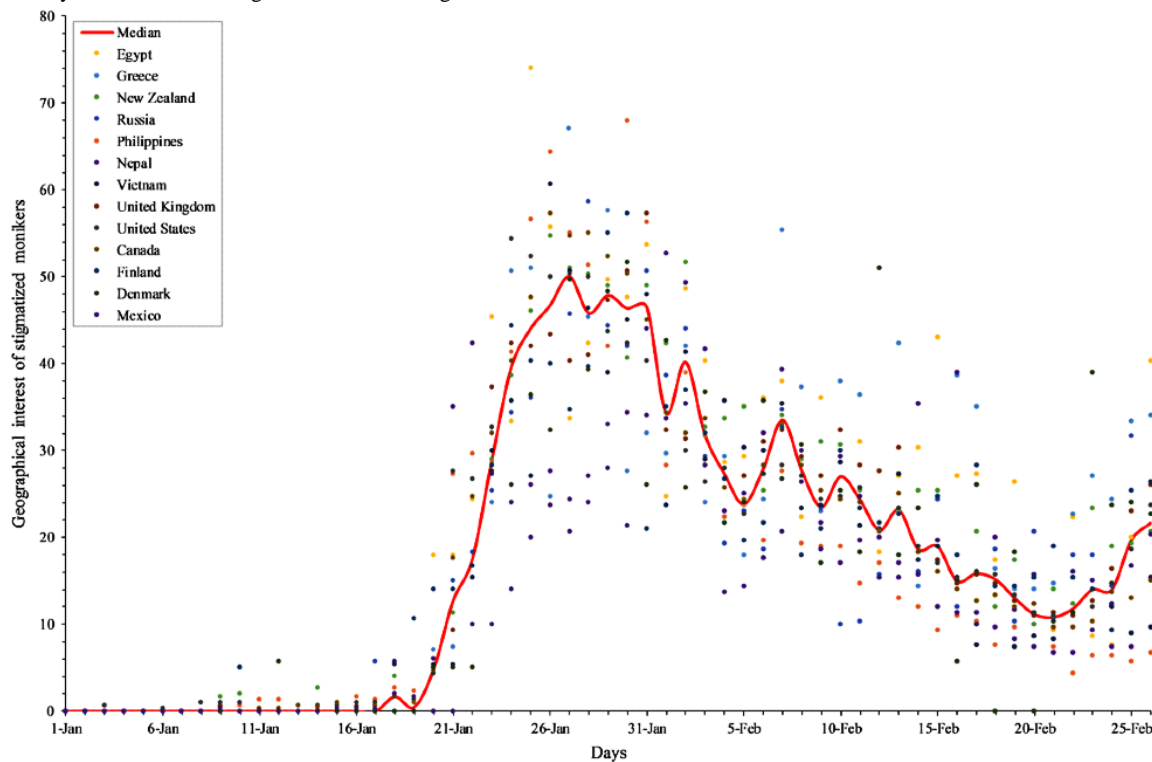
Figure 5. Geographical map of the COVID-19 infodemic (February 29, 2020). WHO: World Health Organization.
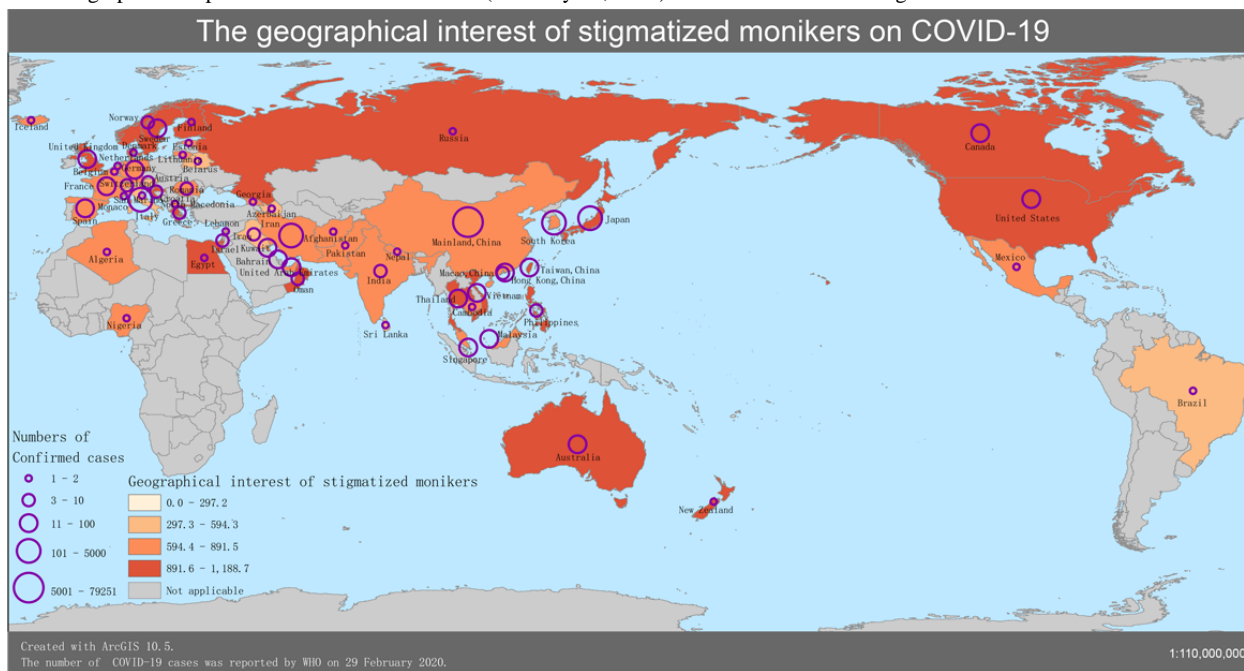
What are the plausible reasons behind such collective perceptual biases? Demographically, according to a Pew Research Center survey [67], a median of 40% of the surveyed countries had a positive view of China, compared with a median of 41% who had an unfavorable opinion. However, recognizing that COVID-19 has a potential public health impact, people are asking existential questions, making them vulnerable to the surfeit of information contagion from the outside world. When emerging cases were reported in their country, infodemics about the cause of the epidemic began, and nothing seemed certain or obviously right. As a case in point, the Bill & Melinda Gates Foundation, an American private foundation that has spent billions on global health care, have been suspiciously accused of manufacturing COVID-19 as biological warfare with the CIA to "wage economic war on China." However, this was evidently not plausible. Such an infodemic campaign reminds us that authoritative organizations should work together with each other and cultivate a well-trained team of professionals to mediate infodemic risks.

From the beginning of the COVID-19 epidemic, people in most Asia-Pacific countries, "where many more name China as a top threat" [67], prioritized relations with China to jointly fight COVID-19, rather than use malicious discrimination against China. Regrettably, some individuals and media outlets have been committed to showing a negative image of China by promoting unfounded conspiracy theories such as a nonnatural origin of COVID-19 [68], a coronavirus that was made in China (even including desecration of the Chinese national flag), "China is the real Sick Man of Asia" [69], and China's Chernobyl moment. Some instigators have made open apologies for spreading these rumors [14], but others are intent on whitewashing their words under the guise of freedom of speech.

Such voices do nothing but breed the pathogen of fear, panic, prejudice, disgust, xenophobia, and racism [9,14,70,71]. Undoubtedly, they have been met overwhelmingly with harsh criticism. On February 8, 2020, the *Lancet* published a statement

in solidarity with Chinese professionals in combating the novel coronavirus outbreak and called for fighting against the infodemics [68,72]. Later, more public health scientists have endorsed this statement.

Collective Perceptual Biases in the Scientific Community

Given that multifarious stigmatized monikers have become dominant in the public, what is being used in the scientific sphere? Admittedly, the plethora of papers on the pandemic have somewhat aggravated the collective perceptual biases, whether intentional or unintentional [73-75]. It is critical to have individuals who are well versed in naming conventions collaborate directly with researchers on a regular basis. Unfortunately, before the antidotes for the infodemic (ie, proper names) find their way into the public mind, debate on interim solutions has been ongoing (Multimedia Appendix 3).

On January 12, 2020, the WHO provisionally named the 2019 novel coronavirus disease "2019-nCoV." China's National Health Commission (CNHC) decided to temporarily call the disease "Novel Coronavirus Pneumonia" or "NCP" on February 7 (Figure 2). The CNHC's official name has invoked intensive arguments outside as well as inside the scientific community. First, Chinese scientists are divided on that official name. Supporters say the descriptive name follows typical classification practices, whereas opponents claim that it could be easily misunderstood and abused to sow the seeds for panic. Second, the word *novel* is confusing because neither the disease nor the host can be used to reliably determine the virus's novelty. Arguably, high mutation and gene recombination rates make this type of virus ideal for pathogen evolution [76]. Once viral mutation happens, it will no longer be *novel*.

Before that, the 2019 novel coronavirus was designated as "WH-Human-1 coronavirus" ("Wuhan-Human-1 coronavirus") by a group of scientists in *Nature* on February 3, 2020 [77]. In the same vein, on February 11, another name, "HARS-CoV"

(Han acute respiratory syndrome coronavirus) with *Han* standing for *Wuhan* in Chinese, was proposed in *The Lancet* (of note, some of the coauthors are members of the WHO IHR Emergency Committee) [15]. Obviously, such practices are against the naming principles of the WHO—geographic locations should be avoided in virus or disease names, and the proper names should be short and easy to pronounce [78]. Such names might provoke unintended negative impacts by stigmatizing Wuhan citizens and Chinese people. Those flawed notions take hold and should be duly corrected, as well as other similar paradigms (Table 1).

Table 1. Chronological list of published articles with multifarious proposed names.

Date (2020)	Article[a]	Proposed name
January 18	Cheng et al [79]	Wuhan coronavirus pneumonia
January 20	Parry [80]	China coronavirus
January 21	*Nature* [81]	Wuhan virus
January 22	Callaway and Cyranoski [82]	China coronavirus
January 22	Liu and Saif [83]	Wuhan coronavirus
January 23	Callaway and Cyranoski [84]	China virus
January 24	Mahase [85]	China coronavirus
January 28	Mahase [86]	China coronavirus
January 29	Parry [87]	China coronavirus
January 31	Callaway [88]	China coronavirus
January 31	Mahase [89]	China coronavirus
January 31	Bassetti et al [90]	Novel Chinese coronavirus
January 31	Ralph et al [47]	Wuhan virus
February 3	Wu et al [77]	WH-Human-1 Coronavirus[b]
February 4	Parry [91]	China coronavirus
February 5	Jiang et al [92]	PARS-CoV[c]
February 7	Cyranoski [93]	China coronavirus
February 11	Wang et al [15]	HARS-CoV[d]
February 11	Coronaviridae Study Group of the International Committee on Taxonomy of Viruses [38]	SARS-CoV-2
February 12	Zhou et al [94]	Wuhan novel coronavirus
February 14	Jiang and Shi [95]	TARS-CoV[e]
February 19	Jiang et al [96]	HCoV-19[f]
February 19	Goh et al [97]	Wuhan-2019-nCoV
February 19	Kooraki et al [98]	NCIP[g]
February 26	Xia et al [99]	NCP[h]

[a]The metadata of the articles were retrieved from PubMed as of February 26, 2020.

[b]WH-Human-1 Coronavirus: Wuhan-Human-1 coronavirus.

[c]PARS-CoV: pneumonia acute respiratory syndrome coronavirus.

[d]HARS-CoV: *Han* acute respiratory syndrome coronavirus.

[e]TARS-CoV: transmissible acute respiratory syndrome coronavirus.

[f]HCoV-19: human coronavirus 2019.

[g]NCIP: novel coronavirus-infected pneumonia.

[h]NCP: novel coronavirus pneumonia.

In response to such concerns, on February 11, 2020, the WHO officially renamed "2019-nCoV" as "COVID-19," with *CO* meaning *corona*, *VI* for *virus*, *D* for *disease*, and *19* referring to 2019. This generic descriptive reassignment offers an overdue correction to those strongly held but flawed notions, with the hope of minimizing stigma. Coinciding with the WHO's latest announcement, in a bioRxiv preprint [100], a new name "Severe Acute Respiratory Syndrome coronavirus 2" or "SARS-CoV-2" was penned by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses (ICTV-CSG) on the same

day. ICTV-CSG explains that this designation highlights the new strain's similarity to SARS-CoV [38]. It is unclear whether this proposal name will be approved by the next plenary meeting of ICTV.

The WHO and some prominent virologists are far less skewed toward SARS-CoV-2, the nomenclature endorsed by ICTV-CSG [96,101]. Outside the academic-industrial sphere, people also argued against this proposed name. Although it seems natural for ICTV-CSG to add a numeral *2* behind "SARS-CoV" to signify their relation, many prominent scientists have scrambled to refute this claim. To the untrained eye, the hasty designation may mislead the public to perceive a more severe virus strain as a direct descendant of SARS-CoV, rather than just a close affinity for the causative agent of another major viral outbreak in China in 2002 and 2003. Before that, on February 5, 2020, Jiang and colleagues [92] proposed another name, "Pneumonia Acute Respiratory Syndrome Coronavirus" ("PARS-CoV") in *Cellular & Molecular Immunology*. By the same token, this assignment also intends to retain equivalent terminology to SARS-CoV. Nonetheless, only 2 weeks later, without mention of their earlier similar formulations [92,95], they reintroduced the third name "HCoV-19" ("Human coronavirus 2019") in *The Lancet* [96], objecting to the usage of SARS-CoV-2.

The looming worry is that the public are susceptible to SARS-CoV [25], which evokes the memory of the higher case fatality ratio. On February 9, 2020, Chen Huan-chun, a Chinese academician and virologist, made a public apology for mistakenly saying 2019-nCoV is SARS-CoV, which had struck a nerve and aroused great consternation in the Chinese public.

Mitigating infodemic risks by making informed and judicious choices is a catch-22 for authoritative organizations. It is necessary to punctuate heuristic cautions and continuous introspection of previous multifarious names [18,25,78,101], which is a requisite bedrock of such scientific efforts. Recently, global profusion of candidates has been discussed inside the scientific community, as well as on social media (eg, transmissible acute respiratory syndrome coronavirus [TARS-CoV] [95] and contagious acute respiratory syndrome coronavirus [CARS-CoV]). Whatever merits and demerits each term has, some of them should be fairly recognized with plausible reasons. Authorities should have an open mind to the modest introspections and rededications of such collective efforts. On February 22, 2020, CNHC officially renamed the temporary English name "NCP" as "COVID-19," with the hope of standing with the WHO and further discouraging the use of stigmatized titles [102].

Combating the COVID-19 Infodemic: Same Mission, Diverse Voices

As the COVID-19 epidemic spreads, so does the information epidemic. The COVID-19 infodemic has introduced a new round of challenges for crisis communication, just as Dr Tedros Adhanom Ghebreyesus, the Director-General of the WHO, remarked at the Munich Security Conference on February 15, 2020: "We're not just fighting an epidemic; we're fighting an infodemic." Infoveillance is an effective strategy against infodemics [34,35,103]. Unfortunately, with the same mission of corroborating reliable information and keeping people informed, different practitioners are upholding diverse voices in the campaigns against the ongoing information epidemic. As can be seen from Figure 6, Google Trends showed that the interest of the portmanteau words "infodemic," "disinfodemic," and "misinfodemic" from January 1, 2020, to July 15, 2020. The discourse system to address the present challenge is divided into three camps: the *infodemic* campaign endorsed by the WHO partnered with internet giants worldwide, the *disinfodemic* campaign backed by organizations led by the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the *misinfodemic* campaign supported by other practitioners. Although the *infodemic* campaign is dominating the fray, most people are currently more interested in what is going on in the real world but are curious about what an infodemic is. The detailed descriptions of the code scheme of combating the COVID-19 infodemic in this study can be found in Multimedia Appendix 3.

In 2002, Eysenbach [27, 103] coined the portmanteau "infodemiology" (a novel transdisciplinary science to unravel the complex propagation patterns of misinformation and public health relevant information) along with the portmanteau "infoveillance" (a type of syndromic surveillance that uses online content). On February 2, 2020, the WHO adopted the term "infodemic" as an "overabundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it" [1]. In the aftermath of the online technical consultation on the COVID-19 infodemic [1], the WHO crystallized an evidence-based framework to underpin infodemic management interventions [1, 2, 104]. In the disinfodemic campaign, Posetti and Bontcheva [105] proposed the neologism "disinfodemic" (a blend of dis-, information, and epidemic) in the research-based policy briefs of UNESCO, considering its opposite of information [106]. A minority of researchers favor the term "misinfodemic" (a blend of mis-, information, and epidemic) in line with misinformation [107]. In contrast, "infodemic" is a more efficient portmanteau than "disinfodemic" or "misinfodemic" for communicative efficiency determined by shorter orthographic and phonetic length, according to Zipf's [108,109] principle of least effort governing human lexicons.

Figure 6. Illustration on the relative search interest of information contagion from around the world. EPI-WIN: Information Network for Epidemics; WHO: World Health Organization.
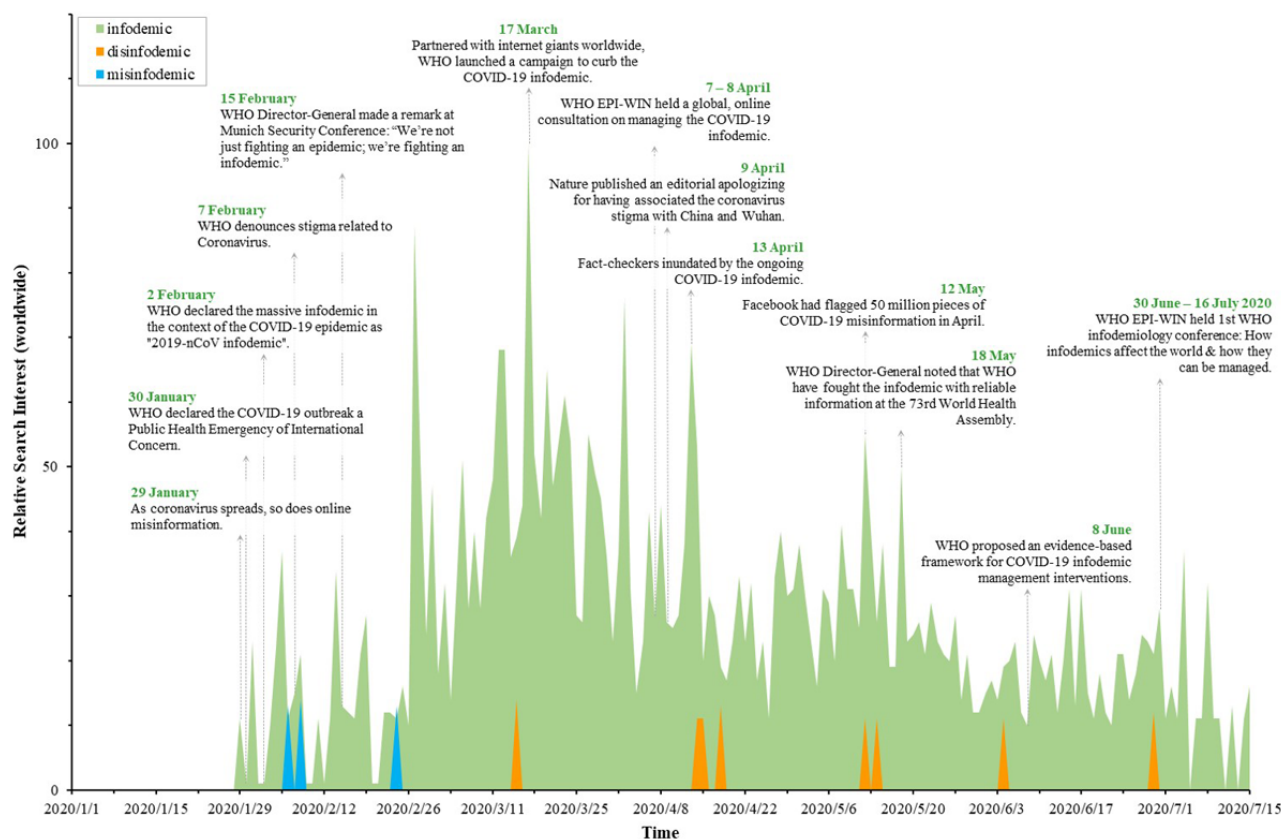
On the other hand, it is critical to engage individuals to fully realize the damaging effect of infodemics and participate in the initiative of disproving COVID-19 myths. With social network giants worldwide, the WHO has been committed to curb the infodemics. For example, as part of the coordinated action, Facebook flagged around 50 million pieces of COVID-19–related content in April 2020. However, according to a Pew Research Center's American News Pathways survey, about three in ten Americans still believe the viral conspiracy theories (COVID-19 was created in a lab and the COVID-19 outbreak was intentionally planned by people in power) [110]. Therefore, drawing on the lessons learned from the contextualized pollution of the media ecosystem, each one of us should contribute to the fight against both societal COVID-19 and information contagion in the most effective way.

Discussion

Principal Findings

With an emphasis on infodemiological analysis and meta-analysis on the COVID-19 epidemic and infodemic, we scrutinized the collective communication behaviors on the internet and pertinent usages in publications in sociocultural paradigms to uncover collective behavioral propensities and consequences.

First, psychologists often make claims about the relatedness between epidemics and panic based on qualitative evidence. The quantitative results reveal that people are invariably vulnerable to panic attacks during episodes of epidemics with an enigmatic nature. People around the world are divided in

their favor of stigmatized monikers because of perceptual bias in the public and scientific communities. People in 13 (22%) out of 58 territories with low cumulative rates had negative behavioral propensities to stigmatized monikers in their daily communications. Perceptual bias in the perception of the natural origin of COVID-19 is part of the reason for specific regions, rather than the degree of infection in their territories.

Second, infodemics follow closely on the heels of every pathogen [5,6,23,24,37], branding discrimination and stoking panic. Official names would duly discourage the spread of regional stigmatization and racial discrimination, and reverse negative perceptual biases and collective behavioral propensities in public engagement.

Third, the coordinated campaign of fighting against the COVID-19 infodemic has called for an approachable uniform voice in line with the same mission, keeping lay audiences informed.

Conclusion

With the benefit of hindsight provided by the Gini coefficient (G), the contextualized results indicate that many stigmatized monikers against China had a higher collective heterogeneity of crowd behavior than the official terms between December 30, 2019, and July 15, 2020. The prognostic significance of information seeking and avoidance is that infodemiological analysis could provide a hallmark reference to reframe extensible discussions on the COVID-19 epidemic and infodemic, as well as substantial patterns of the next infodemic.

At this critical moment, an epoch-making name is expected to be scientifically pithy and socially acceptable, with minimal unintentional negative impacts on nations, economies, and people. This is a positivist doctrine, not merely for naming a virus but for the vitality of science and the promotion of social progress. Obviously, some naming practices went awry, intentionally or not [14]. A learning lesson from the infodemic is the necessity of coming up with guidelines for the adoption of practical principles intended to enhance the possibility for the lessening of stigmatization and discrimination.

Technically, we now see collaborative efforts as a potential way to help strengthen and standardize ongoing international initiatives of the WHO and the ICTV [5,6]. Admittedly, understanding the way naming rules strengthen the integrity and quality of naming practices with the original mission remains nominal rather than substantial [18,25,78,101]. A *Nature* editorial remarked, "As well as naming the illness, the WHO was implicitly sending a reminder to those who had erroneously been associating the virus with Wuhan and with China in their news coverage — including *Nature*. That we did so was an error on our part, for which we take responsibility and apologize" [14]. As another precaution, the word *novel* was recommended by the WHO for "indicating a new pathogen of a previously known type, recognizing that this term will become obsolete if other new pathogens of that type are identified" [21]. However, stakeholders frequently reserve *novel* for indicating new types of viruses, lest this word fundamentally lose its impact without regular amendments.

Acknowledgments

Authors' Contributions

ZH was involved in the conceptual design of the study. ZH, ZY, QL, and AZ performed the metadata analyses. All authors wrote and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The transdisciplinary nature of infodemiology.
[DOCX File , 408 KB - jmir_v22i11e22639_app1.docx]

Multimedia Appendix 2
Diachronic discourse analysis and scientometric analysis.
[DOCX File , 146 KB - jmir_v22i11e22639_app2.docx]

Multimedia Appendix 3
Code schemes.
[DOCX File , 73 KB - jmir_v22i11e22639_app3.docx]

Multimedia Appendix 4
Cumulative rates.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i11e22639_app4.xlsx]

References

1. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
2. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
3. Ratzan SC, Sommariva S, Rauh L. Enhancing global health communication during a crisis: lessons from the COVID-19 pandemic. Public Health Res Pract 2020 Jun 30;30(2):3022010. [doi: 10.17061/phrp3022010] [Medline: 32601655]
4. Novel coronavirus (2019-nCoV) situation report - 13. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf

5. Editorial. Coronavirus misinformation needs researchers to respond. Nature 2020 May;581(7809):355-356. [doi: 10.1038/d41586-020-01550-y] [Medline: 32461657]

6. Ball P, Maxmen A. The epic battle against coronavirus misinformation and conspiracy theories. Nature 2020 May;581(7809):371-374. [doi: 10.1038/d41586-020-01452-z] [Medline: 32461658]

7. Shimizu K. 2019-nCoV, fake news, and racism. Lancet 2020 Feb 29;395(10225):685-686 [FREE Full text] [doi: 10.1016/S0140-6736(20)30357-3] [Medline: 32059801]

8. Chung RY, Li MM. Anti-Chinese sentiment during the 2019-nCoV outbreak. Lancet 2020 Feb 29;395(10225):686-687 [FREE Full text] [doi: 10.1016/S0140-6736(20)30358-5] [Medline: 32122469]

9. Rzymski P, Nowicki M. Preventing COVID-19 prejudice in academia. Science 2020 Mar 20;367(6484):1313. [doi: 10.1126/science.abb4870] [Medline: 32193314]

10. Karalis Noel T. Conflating culture with COVID-19: xenophobic repercussions of a global pandemic. Soc Sci Humanit Open 2020;2(1):100044. [doi: 10.1016/j.ssaho.2020.100044]

11. He J, He L, Zhou W, Nie X, He M. Discrimination and social exclusion in the outbreak of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2933 [FREE Full text] [doi: 10.3390/ijerph17082933] [Medline: 32340349]

12. Rzymski P, Nowicki M. COVID-19-related prejudice toward Asian medical students: a consequence of SARS-CoV-2 fears in Poland. J Infect Public Health 2020 Jun;13(6):873-876 [FREE Full text] [doi: 10.1016/j.jiph.2020.04.013] [Medline: 32387102]

13. Misra S, Le PD, Goldmann E, Yang LH. Psychological impact of anti-Asian stigma due to the COVID-19 pandemic: a call for research, practice, and policy responses. Psychol Trauma 2020 Jul;12(5):461-464. [doi: 10.1037/tra0000821] [Medline: 32525390]

14. Editorial. Stop the coronavirus stigma now. Nature 2020 Apr;580(7802):165. [doi: 10.1038/d41586-020-01009-0] [Medline: 32265571]

15. Wang L, Anderson DE, Mackenzie JS, Merson MH. From Hendra to Wuhan: what has been learned in responding to emerging zoonotic viruses. Lancet 2020 Feb 22;395(10224):e33-e34 [FREE Full text] [doi: 10.1016/S0140-6736(20)30350-0] [Medline: 32059799]

16. Sharot T, Sunstein CR. How people decide what they want to know. Nat Hum Behav 2020 Jan;4(1):14-19. [doi: 10.1038/s41562-019-0793-1] [Medline: 31932690]

17. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]

18. Enserink M. Infectious diseases. Amid heightened concerns, new name for novel coronavirus emerges. Science 2013 May 10;340(6133):673. [doi: 10.1126/science.340.6133.673] [Medline: 23661733]

19. Goodwin R, Haque S, Neto F, Myers LB. Initial psychological responses to Influenza A, H1N1 ("Swine flu"). BMC Infect Dis 2009 Oct 06;9:166 [FREE Full text] [doi: 10.1186/1471-2334-9-166] [Medline: 19807908]

20. Dhand NK, Hernandez-Jover M, Taylor M, Holyoake P. Public perceptions of the transmission of pandemic influenza A/H1N1 2009 from pigs and pork products in Australia. Prev Vet Med 2011 Feb 01;98(2-3):165-175. [doi: 10.1016/j.prevetmed.2010.11.006] [Medline: 21146879]

21. World Health Organization best practices for the naming of new human infectious diseases. World Health Organization. 2015 May. URL: https://www.who.int/topics/infectious_diseases/naming-new-diseases/en/

22. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]

23. Rothkopf D. When the buzz bites back. The Washington Post. 2003 May 11. URL: http://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html

24. Garcia E, Yactayo S, Nishino K, Millot V, Perea W, Brianda S. Zika virus infection: global update on epidemiology and potentially associated clinical manifestations. Wkly Epidemiol Rec 2016 Feb 19;91(7):73-81 [FREE Full text] [Medline: 26897760]

25. Enserink M. War stories. Science 2013 Mar 15;339(6125):1264-1268. [doi: 10.1126/science.339.6125.1264] [Medline: 23493690]

26. Davis JT, Perra N, Zhang Q, Moreno Y, Vespignani A. Phase transitions in information spreading on structured populations. Nat Phys 2020 Mar 2;16(5):590-596. [doi: 10.1038/s41567-020-0810-3]

27. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

28. Horvitz E, Mulligan D. Policy forum. Data, privacy, and the greater good. Science 2015 Jul 17;349(6245):253-255. [doi: 10.1126/science.aac4520] [Medline: 26185242]

29. 2019 revision of world population prospects. United Nations Population Division, Department of Economic and Social Affairs. 2019. URL: https://population.un.org/wpp/

30. Michel J, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science 2011 Jan 14;331(6014):176-182 [FREE Full text] [doi: 10.1126/science.1199644] [Medline: 21163965]

31. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature 2018 Mar 07;555(7695):175-182. [doi: 10.1038/nature25753] [Medline: 29517004]

32. Fancourt D, Steptoe A, Wright L. The Cummings effect: politics, trust, and behaviours during the COVID-19 pandemic. Lancet 2020 Aug 15;396(10249):464-465. [doi: 10.1016/S0140-6736(20)31690-1] [Medline: 32771083]

33. Scheufele DA, Krause NM. Science audiences, misinformation, and fake news. Proc Natl Acad Sci U S A 2019 Apr 16;116(16):7662-7669 [FREE Full text] [doi: 10.1073/pnas.1805871115] [Medline: 30642953]

34. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]

35. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

36. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Dutilh BE, Harrach B, et al. Binomial nomenclature for virus species: a consultation. Arch Virol 2020 Feb;165(2):519-525 [FREE Full text] [doi: 10.1007/s00705-019-04477-6] [Medline: 31797129]

37. Jones DS. History in a crisis - lessons for COVID-19. N Engl J Med 2020 Apr 30;382(18):1681-1683. [doi: 10.1056/NEJMp2004361] [Medline: 32163699]

38. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]

39. Almeida JD, Berry DM, Cunningham CH, Hamre D, Hofstad MS, Mallucci L, et al. Virology: coronaviruses. Nature 1968 Nov;220(5168):650-650. [doi: 10.1038/220650b0]

40. Fenner F. The nomenclature and classification of viruses the International Committee on Nomenclature of Viruses. Virology 1971 Dec;46(3):979-980. [doi: 10.1016/0042-6822(71)90102-4]

41. Fenner F. The classification and nomenclature of viruses. Summary of results of meetings of the International Committee on Taxonomy of Viruses in Madrid, September 1975. J Gen Virol 1976 Jun;31(3):463-470. [doi: 10.1099/0022-1317-31-3-463] [Medline: 819628]

42. Kupferschmidt K, Cohen J. Can China's COVID-19 strategy work elsewhere? Science 2020 Mar 06;367(6482):1061-1062. [doi: 10.1126/science.367.6482.1061] [Medline: 32139521]

43. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016 Aug;14(8):523-534 [FREE Full text] [doi: 10.1038/nrmicro.2016.81] [Medline: 27344959]

44. Cui J, Li F, Shi Z. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019 Mar;17(3):181-192 [FREE Full text] [doi: 10.1038/s41579-018-0118-9] [Medline: 30531947]

45. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]

46. Wadman M. How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. Science 2020 Apr 18:1. [doi: 10.1126/science.abc3208]

47. Ralph R, Lew J, Zeng T, Francis M, Xue B, Roux M, et al. 2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission, travel-related cases, and vaccine readiness. J Infect Dev Ctries 2020 Jan 31;14(1):3-17 [FREE Full text] [doi: 10.3855/jidc.12425] [Medline: 32088679]

48. Carlos WG, Dela Cruz CS, Cao B, Pasnick S, Jamil S. Novel Wuhan (2019-nCoV) coronavirus. Am J Respir Crit Care Med 2020 Feb 15;201(4):P7-P8. [doi: 10.1164/rccm.2014P7] [Medline: 32004066]

49. Chan JFW, Lau SKP, To KKW, Cheng VCC, Woo PCY, Yuen K. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. Clin Microbiol Rev 2015 Apr;28(2):465-522 [FREE Full text] [doi: 10.1128/CMR.00102-14] [Medline: 25810418]

50. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]

51. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiatry 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/S2215-0366(20)30046-8] [Medline: 32032543]

52. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and human-to-human transmission of a novel coronavirus in vietnam. N Engl J Med 2020 Feb 27;382(9):872-874 [FREE Full text] [doi: 10.1056/NEJMc2001272] [Medline: 31991079]

53. Li J, Li J, Xie X, Cai X, Huang J, Tian X, et al. Game consumption and the 2019 novel coronavirus. Lancet Infect Dis 2020 Mar;20(3):275-276 [FREE Full text] [doi: 10.1016/S1473-3099(20)30063-3] [Medline: 32043979]

54. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020 Feb;91:264-266 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.009] [Medline: 31953166]

55. Su S, Wong G, Shi W, Liu J, Lai AC, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016 Jun;24(6):490-502 [FREE Full text] [doi: 10.1016/j.tim.2016.03.003] [Medline: 27012512]

56. Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen K. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov 2016 May;15(5):327-347 [FREE Full text] [doi: 10.1038/nrd.2015.37] [Medline: 26868298]

57. Peiris JSM, Guan Y, Yuen KY. Severe acute respiratory syndrome. Nat Med 2004 Dec;10(12 Suppl):S88-S97 [FREE Full text] [doi: 10.1038/nm1143] [Medline: 15577937]

58. Normile D. Novel human virus? Pneumonia cases linked to seafood market in China stir concern. Science 2020 Jan 04:1. [doi: 10.1126/science.aba7672]

59. Normile D. China reports more than 200 infections with new coronavirus from Wuhan. Science 2020 Jan 21:1. [doi: 10.1126/science.aba9913]

60. Horton R. Offline: 2019-nCoV outbreak-early lessons. Lancet 2020 Feb 01;395(10221):322 [FREE Full text] [doi: 10.1016/S0140-6736(20)30212-9] [Medline: 32007152]

61. Kock R, Karesh W, Veas F, Velavan T, Simons D, Mboera L, et al. 2019-nCoV in context: lessons learned? Lancet Planet Health 2020 Mar;4(3):e87-e88 [FREE Full text] [doi: 10.1016/S2542-5196(20)30035-8] [Medline: 32035507]

62. Lam TT, Jia N, Zhang Y, Shum MH, Jiang J, Zhu H, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 2020 Jul;583(7815):282-285. [doi: 10.1038/s41586-020-2169-0] [Medline: 32218527]

63. Editorial. Exploring wet markets. Nat Food 2020 May 19;1(5):241-241. [doi: 10.1038/s43016-020-0090-1]

64. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]

65. The Lancet. COVID-19: too little, too late? Lancet 2020 Mar 07;395(10226):755 [FREE Full text] [doi: 10.1016/S0140-6736(20)30522-5] [Medline: 32145772]

66. Cowper A. Covid-19: are we getting the communications right? BMJ 2020 Mar 06;368:m919. [doi: 10.1136/bmj.m919] [Medline: 32144115]

67. Silver L, Devlin K, Huang C. China's economic growth mostly welcomed in emerging markets, but neighbors wary of its influence. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/12/05/chinas-economic-growth-mostly-welcomed-in-emerging-markets-but-neighbors-wary-of-its-influence/

68. Calisher C, Carroll D, Colwell R, Corley RB, Daszak P, Drosten C, et al. Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19. Lancet 2020 Mar 07;395(10226):e42-e43 [FREE Full text] [doi: 10.1016/S0140-6736(20)30418-9] [Medline: 32087122]

69. Mead WR. China is the real sick man of Asia. Wall Street Journal. 2020. URL: https://www.wsj.com/articles/china-is-the-real-sick-man-of-asia-11580773677

70. Devakumar D, Selvarajah S, Shannon G, Muraya K, Lasoye S, Corona S, et al. Racism, the public health crisis we can no longer ignore. Lancet 2020 Jun 27;395(10242):e112-e113 [FREE Full text] [doi: 10.1016/S0140-6736(20)31371-4] [Medline: 32534630]

71. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar 21;395(10228):942-943 [FREE Full text] [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]

72. The Lancet. COVID-19: fighting panic with information. Lancet 2020 Feb 22;395(10224):537 [FREE Full text] [doi: 10.1016/S0140-6736(20)30379-2] [Medline: 32087777]

73. London AJ, Kimmelman J. Against pandemic research exceptionalism. Science 2020 May 01;368(6490):476-477. [doi: 10.1126/science.abc1731] [Medline: 32327600]

74. Brainard J. New tools aim to tame pandemic paper tsunami. Science 2020 May 29;368(6494):924-925. [doi: 10.1126/science.368.6494.924] [Medline: 32467369]

75. Casigliani V, De Nard F, De Vita E, Arzilli G, Grosso FM, Quattrone F, et al. Too much information, too little evidence: is waste in research fuelling the covid-19 infodemic? BMJ 2020 Jul 06;370:m2672. [doi: 10.1136/bmj.m2672] [Medline: 32631897]

76. Morens DM, Daszak P, Taubenberger JK. Escaping Pandora's box - another novel coronavirus. N Engl J Med 2020 Apr 02;382(14):1293-1295. [doi: 10.1056/NEJMp2002106] [Medline: 32101660]

77. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Mar;579(7798):265-269 [FREE Full text] [doi: 10.1038/s41586-020-2008-3] [Medline: 32015508]

78. Fukuda K, Wang R, Vallat B. Naming diseases: first do no harm. Science 2015 May 08;348(6235):643. [doi: 10.1126/science.348.6235.643] [Medline: 25954000]

79. Cheng VCC, Wong SC, To KKW, Ho PL, Yuen KY. Preparedness and proactive infection control measures against the emerging novel coronavirus in China. J Hosp Infect 2020 Mar;104(3):254-255 [FREE Full text] [doi: 10.1016/j.jhin.2020.01.010] [Medline: 31962139]

80. Parry J. China coronavirus: cases surge as official admits human to human transmission. BMJ 2020 Jan 20;368:m236. [doi: 10.1136/bmj.m236] [Medline: 31959587]

81. Editorial. Stop the Wuhan virus. Nature 2020 Jan;577(7791):450. [doi: 10.1038/d41586-020-00153-x] [Medline: 31965110]

82. Callaway E, Cyranoski D. China coronavirus: six questions scientists are asking. Nature 2020 Jan;577(7792):605-607. [doi: 10.1038/d41586-020-00166-6] [Medline: 31992880]

83. Liu SL, Saif L. Emerging viruses without borders: the Wuhan coronavirus. Viruses 2020 Jan 22;12(2):130 [FREE Full text] [doi: 10.3390/v12020130] [Medline: 31979013]

84. Callaway E, Cyranoski D. Why snakes probably aren't spreading the new China virus. Nature 2020 Jan 23:1. [doi: 10.1038/d41586-020-00180-8]

85. Mahase E. China coronavirus: what do we know so far? BMJ 2020 Jan 24;368:m308. [doi: 10.1136/bmj.m308] [Medline: 31980434]

86. Mahase E. China coronavirus: mild but infectious cases may make it hard to control outbreak, report warns. BMJ 2020 Jan 28;368:m325. [doi: 10.1136/bmj.m325] [Medline: 31992570]

87. Parry J. China coronavirus: partial border closures into Hong Kong are not enough, say doctors. BMJ 2020 Jan 29;368:m349. [doi: 10.1136/bmj.m349] [Medline: 31996345]

88. Callaway E. China coronavirus: labs worldwide scramble to analyse live samples. Nature 2020 Feb;578(7793):16. [doi: 10.1038/d41586-020-00262-7] [Medline: 32020116]

89. Mahase E. China coronavirus: WHO declares international emergency as death toll exceeds 200. BMJ 2020 Jan 31;368:m408. [doi: 10.1136/bmj.m408] [Medline: 32005727]

90. Bassetti M, Vena A, Giacobbe DR. The novel Chinese coronavirus (2019-nCoV) infections: challenges for fighting the storm. Eur J Clin Invest 2020 Mar;50(3):e13209 [FREE Full text] [doi: 10.1111/eci.13209] [Medline: 32003000]

91. Parry J. China coronavirus: Hong Kong health staff strike to demand border closure as city records first death. BMJ 2020 Feb 04;368:m454. [doi: 10.1136/bmj.m454] [Medline: 32019754]

92. Jiang S, Xia S, Ying T, Lu L. A novel coronavirus (2019-nCoV) causing pneumonia-associated respiratory syndrome. Cell Mol Immunol 2020 May;17(5):554 [FREE Full text] [doi: 10.1038/s41423-020-0372-4] [Medline: 32024976]

93. Cyranoski D. Did pangolins spread the China coronavirus to people? Nature 2020 Feb 7:1. [doi: 10.1038/d41586-020-00364-2]

94. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J Evid Based Med 2020 Feb;13(1):3-7 [FREE Full text] [doi: 10.1111/jebm.12376] [Medline: 32048815]

95. Jiang S, Shi Z. The first disease X is caused by a highly transmissible acute respiratory syndrome coronavirus. Virol Sin 2020 Jun;35(3):263-265 [FREE Full text] [doi: 10.1007/s12250-020-00206-5] [Medline: 32060789]

96. Jiang S, Shi Z, Shu Y, Song J, Gao GF, Tan W, et al. A distinct name is needed for the new coronavirus. Lancet 2020 Mar 21;395(10228):949 [FREE Full text] [doi: 10.1016/S0140-6736(20)30419-0] [Medline: 32087125]

97. Goh GKM, Dunker AK, Foster JA, Uversky VN. Rigidity of the outer shell predicted by a protein intrinsic disorder model sheds light on the COVID-19 (Wuhan-2019-nCoV) infectivity. Biomolecules 2020 Feb 19;10(2):331 [FREE Full text] [doi: 10.3390/biom10020331] [Medline: 32092911]

98. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) outbreak: what the department of radiology should know. J Am Coll Radiol 2020 Apr;17(4):447-451 [FREE Full text] [doi: 10.1016/j.jacr.2020.02.008] [Medline: 32092296]

99. Xia J, Tong J, Liu M, Shen Y, Guo D. Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. J Med Virol 2020 Jun;92(6):589-594 [FREE Full text] [doi: 10.1002/jmv.25725] [Medline: 32100876]

100. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses – a statement of the Coronavirus Study Group. bioRxiv. Preprint posted online February 11, 2020 2020. [doi: 10.1101/2020.02.07.937862]

101. Enserink M. Update: 'A bit chaotic.' Christening of new coronavirus and its disease name create confusion. Science 2020 Feb 13:1. [doi: 10.1126/science.abb2806]

102. Zhihao Z. China adopts COVID-19 as official English name for disease. China Daily. 2020. URL: https://www.chinadaily.com.cn/a/202002/22/WS5e506f18a31012821727970f.html [accessed 2020-02-22]

103. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]

104. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]

105. Posetti J, Bontcheva K. Disinfodemic: dissecting responses to COVID-19 disinformation. UNESCO. 2020. URL: https://en.unesco.org/sites/default/files/disinfodemic_dissecting_responses_covid19_disinformation.pdf

106. Callaway E, Ledford H, Mallapaty S. Six months of coronavirus: the mysteries scientists are still racing to solve. Nature 2020 Jul;583(7815):178-179. [doi: 10.1038/d41586-020-01989-z] [Medline: 32620885]

107. Sahoo SS, Sahu DP, Kankaria A. Mis-infodemic: the Achilles' heel in combating the COVID-19 pandemic in an Indian perspective. Monaldi Arch Chest Dis 2020 Jun 03;90(2):1. [doi: 10.4081/monaldi.2020.1405] [Medline: 32498502]

108. Zipf GK. The Psycho-Biology of Language: An Introduction to Dynamic Philology. 1st ed. Boston, MA: Houghton Mifflin Company; 1935.

109. Zipf GK. Human Behavior and the Principle of Least Effort: an Introduction to Human Ecology. Cambridge, MA: Addison-Wesley Press; 1949.

110. Mitchell A, Jurkowitz M, Oliphant JB, Shearer E. Three months in, many Americans see exaggeration, conspiracy theories and partisanship in COVID-19 news. Pew Research Center. 2020. URL: https://www.journalism.org/wp-content/uploads/sites/8/2020/06/PJ_2020.06.29_COVID-19-News-Coverage_FINAL.pdf

Abbreviations

CARS-CoV: contagious acute respiratory syndrome coronavirus
CNHC: China's National Health Commission
GBNC: Google Books Ngram Corpus
GTI: Google Trends Index
HARS-CoV: *Han* acute respiratory syndrome coronavirus
HCoV: human coronavirus
HCoV-19: human coronavirus 2019
ICNV: International Committee on Nomenclature of Viruses
ICTV: International Committee on Virus Taxonomy
ICTV-CSG: Coronavirus Study Group of the International Committee on Taxonomy of Viruses
IHR: International Health Regulations
MERS-CoV: Middle East respiratory syndrome–related coronavirus
NCP: novel coronavirus pneumonia
PARS-CoV: pneumonia acute respiratory syndrome coronavirus
PHEIC: Public Health Emergency of International Concern
SARS-CoV: severe acute respiratory syndrome–related coronavirus
TARS-CoV: transmissible acute respiratory syndrome coronavirus
UNESCO: United Nations Educational, Scientific and Cultural Organization
WH-Human-1 coronavirus: Wuhan-Human-1 coronavirus
WHO: World Health Organization
WoS: Web of Science
2019-nCoV: novel coronavirus

XSL•FO
RenderX

Original Paper

Co-Design to Support the Development of Inclusive eHealth Tools for Caregivers of Functionally Dependent Older Persons: Social Justice Design

Karine Latulippe[1], PhD; Christine Hamel[1], PhD; Dominique Giroux[2,3], PhD

[1]Department of Studies of Teaching and Learning, Laval University, Québec, QC, Canada

[2]Département de réadaptation, Faculté de médecine, Laval University, Québec, QC, Canada

[3]Centre d'Excellence du Vieillissement de Québec, Chu de Québec, Québec, QC, Canada

Corresponding Author:

Karine Latulippe, PhD
Department of Studies of Teaching and Learning
Laval University
2320, rue des Bibliothèques
Québec, QC, G1V 0A6
Canada
Phone: 1 418 435 8541
Email: karine.latulippe.3@ulaval.ca

Abstract

Background: eHealth can help reduce social health inequalities (SHIs) as much as it can exacerbate them. Taking a co-design approach to the development of eHealth tools has the potential to ensure that these tools are inclusive. Although the importance of involving future users in the development of eHealth tools to reduce SHIs is highlighted in the scientific literature, the challenges associated with their participation question the benefits of this involvement as co-designers in a real-world context.

Objective: On the basis of Amartya Sen's theoretical framework of social justice, the aim of this study is to explore how co-design can support the development of an inclusive eHealth tool for caregivers of functionally dependent older persons.

Methods: This study is based on a social justice design and participant observation as part of a large-scale research project funded by the Ministry of Families as part of the Age-Friendly Quebec Program (Québec Ami des Aînés). The analysis was based on the method developed by Miles and Huberman and on Paillé's analytical questioning method.

Results: A total of 78 people participated in 11 co-design sessions in 11 Quebec regions. A total of 24 preparatory meetings and 11 debriefing sessions were required to complete this process. Co-designers participated in the creation of a prototype to support the search for formal services for caregivers. The majority of participants (except for 2) significantly contributed to the tool's designing. They also incorporated conversion factors to ensure the inclusiveness of the eHealth tool, such as an adequate level of digital literacy and respect for the caregiver's help-seeking process. In the course of the experiment, the research team's position regarding its role in co-design evolved from a neutral posture and promoting co-designer participation to one that was more pragmatic.

Conclusions: The use of co-design involving participants at risk of SHIs does not guarantee innovation, but it does guarantee that the tool developed will comply with their process of help-seeking and their literacy level. Time issues interfere with efforts to carry out a democratic process in its ideal form. It would be useful to single out some key issues to guide researchers on what should be addressed in co-design discussions and what can be left out to make optimal use of this approach in a real-world context.

XSL•FO
RenderX

Introduction

Background

The growing emergence of eHealth tools in recent years clearly demonstrates the interest that both individuals and health institutions take in them [1]. Since the onset of the COVID-19 pandemic, this interest has increased. eHealth—the digital tools designed to improve health—enables people to access health information anytime, anywhere [2] and holds the promise of improving access to health care and services [3-6]. However, a literature review has brought to light a digital divide, that is, the unequal opportunity to use eHealth among population groups, which further exacerbates the existing social health inequalities (SHIs) [7]. SHIs are differences between population groups in their health status, including life expectancy and comorbidities, because of modifiable social factors [8]. Already present in our communities, SHI is accentuated by eHealth because the people who have difficulty using eHealth tools are the same people who are at risk of SHIs [7]. Nevertheless, promising solutions have been identified to reduce the risk that eHealth may exacerbate SHI and reap its benefits instead [7]. One of these solutions is to involve people at risk of SHIs in the development of eHealth tools [9,10].

Participatory Research and Co-Design

This participation can be achieved in several ways. Participatory research can be broken down to more specific approaches, including community-based participatory research, action research, participative evaluation, emancipatory evaluation, and co-design. According to the various studies that claim to be participatory, the roles that future users (FUs) play vary from consultant to research team member [11,12]. However, in participatory research, there is an expectation of research co-governance by academic researchers and the target population (eg, FUs of knowledge, services, innovation, research findings, etc) [13]. In research design, and in the health field in particular, there is an increasing amount of research on co-creation or co-design of services, programs, or tools together with FUs [14]. Co-design implies that FUs are considered co-designers in the same way as the research team [15]. This involves cooperation between various experts, such as researchers, designers, or developers, and FUs (experts in their own experience). Particular attention is paid to the participation of FUs in the design process and the centralization of their experiences [16]. It is important to note that FUs are involved in all stages of the creative process and that they are part of the team. This is different from other processes where there is a back and forth between the research team that creates the tool and the collection of data from FUs (interviews or observations) [15]. For the purpose of comparison, interviewing FUs allows designers to listen to and interpret what others *say* and observation focuses on seeing what others are *doing* and how they use products or services, whereas co-design enables FUs to jointly *explore* and *create* solutions [16,17].

Arguments in Favor of Using Participatory Approaches

Researchers and designers advocating participatory approaches such as co-design argue, in particular, that the involvement of FUs in the development of a digital tool makes it possible to determine the interests and capacities of FUs and their attitudes, beliefs, values, and expectations [10,18-20], which increases the chances of developing a universally accessible tool [10]. Universal access includes equipment, internet connections, skills development, ongoing technical support, and appropriate content [10], thereby precluding SHI exacerbation through eHealth. Just as patient partners support the relevance and utility of ongoing research, digital patients (e-patients) can contribute to the development of an equally useful and relevant eHealth tool [21]. Involving patients or FUs in research and innovation also helps to redirect and improve the research project, reduce clinical uncertainty, and accelerate the adoption of significant and highly promising results, with the ultimate goal of improving the health care experience and health outcomes [22,23]. In addition, it is possible that FU participation may prove to be a beneficial element in service design, leading to innovation [24]. Finally, the participation of FUs promotes mutual understanding of the context in which the eHealth tool is used [25]. Advocates of participatory and co-design research stress the fact that these approaches increase the cultural and logistical relevance of action in the community, support empowerment, and promote knowledge transfer [26]. It is also argued that these approaches highlight facilitators and social, political, and economic barriers to the knowledge and resources needed for health [13]. At the individual level, people who engage in participatory research processes perceive beneficial effects on their physical and psychological health, self-confidence, self-esteem, empowerment, and social relationships [27]. Using traditional data collection methods such as interviews and observation to grasp multiple facets of the FU's situation would require asking the right questions, conducting the interview or observation at the right time, and interpreting findings in the right way. Co-design enables co-designers to incorporate all this information into the development of the tool [15]. All these arguments support the idea that co-design can contribute to the reduction of SHIs through the opportunity to make the developed tool relevant and accessible to all FUs.

Arguments That Add Nuance to the Use of Participatory Approaches

Researchers and designers have raised important counter-arguments to nuance the potential benefits of co-design. Among other things, people may not be aware of their needs, may be unable to express their needs, or may be unwilling to discuss them in a group setting [28]. In addition, as this type of approach often involves a small number of people, the technology developed may be overspecialized and relevant only to a few people [28]. Furthermore, involving users is not an easy task for researchers and designers in terms of time, energy, and competence. The added value of this type of approach (the benefits of using this approach vs the time and energy required) must be clear and explicit, which is not the case at the moment [13,28]. In addition, there may be disagreements within the group that lead to negotiations and compromises that result in inconclusive outcomes [19]. Moreover, stakeholders who are skeptical about applying this type of approach to digital technologies point out the difficulties for people with poor technological skills to contribute significantly to the design of the tool [15]. The well-known statement attributed to Henry

Ford is relevant in this context: if the public had been consulted to improve the car, they would have added horses. At the individual level, people who engage in a participatory research process may experience physical and psychological exhaustion, stress, and financial loss [27]. Disappointment with the results of the process is also a possible outcome [27]. In the end, it is difficult to understand the mechanism underlying the beneficial effects of participatory research and co-design [29-31] and, more specifically, to determine the potential offered by co-design for the development of eHealth tools that would help reduce rather than exacerbate SHIs.

Theoretical Framework

The development of eHealth tools with a view to reducing rather than exacerbating SHIs is a social justice issue. In this sense, the theoretical framework of social justice by Sen [32] and his capability approach shed light on the potential of co-design to reduce SHIs in eHealth at 2 levels: as a democratic process and as a conversion factor.

Co-Design as a Democratic Process

At a certain point, capabilities, defined as the real freedom that individuals have to *be* or *do* what they have reason to value [33], come to clash, namely, between individual and collective freedoms (or preferences). For example, the choice of having a free public health system, which theoretically allows everyone the freedom to take care of their health, implies a societal choice to contribute financially to support it. For Sen [32], only individual preferences that pass the democratic debate test deserve to be supported by public action insofar as this debate leaves as much space as possible for the individuals concerned [34]. Although Sen addresses this democratic process at the population level, his capability approach provides a relevant interpretive framework for the co-design process. Bonvin and Farvaque [34] explain that from Sen's perspective, the democratic process has 3 values: (1) an intrinsic value that implies that the possibility of participating in a public debate is a fundamental capability, regardless of the final result; (2) an instrumental value where people can express their views and the latter are integrated into decision making; and (3) a constructive value that reveals that each way of considering a problem is a social construct and that it is important that the people concerned are involved in the process. This suggests the guidelines for the development of an eHealth tool in a democratic way: (1) supporting the participation of people at risk of SHIs in the co-design process, both through the opportunity to attend co-design sessions (CoDs) and through access to free speech and unfiltered information; (2) sharing decision-making power with people at risk of SHIs, including efforts to neutralize inequalities [34].

Co-Design as a Conversion Factor

Conversion factors are various personal, social, and environmental characteristics that positively or negatively affect an individual's ability to convert their resources and formal rights into effective functioning [35]. In this study, conversion factors are facilitators and barriers to individuals' use of eHealth tools to take care of their health. The participation of FUs appears to be a facilitator for the co-development of tools for people at risk of SHIs to the degree that it can make these tools more accessible and tailored to their needs.

Objectives of This Paper

The context of this paper is a study aimed at exploring 7 conversion factors that must be considered in the development of eHealth tools, with a view to reducing SHIs. This paper focuses specifically on one of these factors, the participation of people at risk of SHIs in the co-design of an eHealth tool. The objective of this paper is to explore the ways in which co-design can support the development of an inclusive eHealth tool for caregivers of functionally dependent older persons. This paper is the third in a series of papers included in a thesis on the reduction of SHIs in an eHealth context [7,36]. The methodology behind this thesis is detailed in the second paper of this series [36] and is briefly summarized here.

Methods

Study Design

This study is part of a larger project titled, "Better meeting the needs of caregivers in providing safe home care for the functionally impaired older persons," that the research team informally refers to as "the QADA project" in recognition of the fact that it is funded by the Ministry of Families as part of the Québec Ami des Aînés (QADA; Age-Friendly Quebec Program). The project is led by a group of researchers whose intent is to include the social justice perspective in their project (more details are given in the protocol of this project) [37]. The purpose of the QADA project is to develop an eHealth tool to facilitate the process of help-seeking for the caregiver of functionally dependent older persons. The QADA project is based on a co-design approach and thus allows us to achieve our objective.

This study is qualitative in nature, with what can be described as a social justice design since the concept of social justice, based on the capabilities approach, is involved in all phases of the study [38]. Using a qualitative approach was essential to describing the mechanism that potentially underlies the effect of co-design on SHIs [39].

Population, Participants, and Selection Criteria

All QADA project co-designers are participants in this study. They fall into 4 categories: caregivers, community workers, health and social service professionals (HSSPs), and research team.

Caregivers of Functionally Dependent Older Persons

This population is at risk of SHIs because of they are much more at risk of developing health problems (physical and psychological) than the general population [40-42]. For the purpose of this project, any person who provides unpaid assistance on a regular (weekly) basis to a functionally dependent older person is considered a caregiver.

Community Workers and Health and Social Service Professionals

Given their proximity to caregivers, the possibility of obtaining an additional perspective and the desire to develop a tool that

is complementary to what exists already, the choice to involve community workers and HSSPs as co-designers was relevant to the QADA project. The condition for participating was that they offer services or interact directly with caregivers of functionally dependent older persons (eg, nurses, home care providers, social workers).

Research Team

The members of the research team are the participants, and this is of key importance in this study, insofar as the integration of conversion factors must rest on an epistemological and methodological choice made by researchers and designers that must be applied in a realistic and concrete way. Their point of view, which will be largely experiential within the QADA project, is therefore crucial for the implementation of the recommendations resulting from this study. The QADA project research team initially consisted of 8 coresearchers. The participation of these individuals varies according to their availability and expertise. The members of the research team are the 4 participants involved in all phases of the project, and they included the QADA project director, an anthropologist and professional researcher, a user experience designer, and the author of this paper—a doctoral candidate in educational technology. The author of this paper was involved in this study as a participant observer [43], that is, the author took part in the preparation of the CoDs by ensuring the participation of FUs, facilitating the CoDs, debriefing CoDs, and developing the prototype from the results of the CoDs. However, the author played the role of observer when listening to recordings and analysis, steps that began once the co-design phase was completed.

Recruitment

Recruitment was performed through the QADA project with a purposive sampling strategy [37]. The home care and support for the autonomy of seniors programs older adult care management of the 11 Integrated Health and Social Service Centres were contacted to recruit HSSPs. Members of community organizations were contacted directly via phone or email. They were asked to publicize our recruitment announcement among caregivers attending their institutions and activities. Finally, recruitment announcements for caregivers were posted to 30 family medicine groups throughout the province. The latter method of recruitment did not work. Caregivers were therefore recruited through the services of HSSPs and community workers, which implies that they were already service users.

The Research Sites

The study took place in 11 Quebec regions (Côte-Nord, Mauricie, Centre-du-Québec, Capitale-Nationale, Chaudière-Appalaches, Montérégie, Bas St-Laurent, Gaspésie, Outaouais, Montreal, and Laval). The location of CoDs varied, depending on availability (eg, municipal or community premises or those connected with the HSSP network). The research team's work sessions (preparation sessions) were held at the research center, sometimes in person, sometimes using Skype (Microsoft)—with members in remote locations. Table 1 shows the number and type of co-designers who took part in each CoD. Co-designers were invited to participate in the CoD held in their region. This meant that the co-designers were not the same from one session to another except for the 3 advisory committee (AC) meetings, where it was hoped that the participants would be the same.

Table 1. Number of preparation sessions required for the research team, the number and type of co-designers at each co-design sessions, and the content covered in co-design sessions.

CoD[a]	Number of preparation sessions required (n=24) by the research team[b]	Number and type of co-designers (N=74+4 research team members)	Content covered in co-design or AC[c] sessions
CoD1	2	2 CWs[d], 2 HSSPs[e], 3 caregivers	Identification of caregivers' needs
CoD2	1	1 CW, 1 HSSP, 4 caregivers	Identification of caregivers' needs
AC1	1	2 CWs, 2 HSSPs, 1 caregiver	Final choice of needs and recommendations for the continuation of co-design
CoD3	1	2 CWs, 2 HSSPs, 2 caregivers	Exploration of existing functionalities that meet needs and identifying gaps
CoD4	2	2 CWs, 2 HSSPs, 1 caregiver	Brainstorming on the functionalities that can address the gaps
CoD5	3	3 CWs, 2 HSSPs, 3 caregivers	Choice of functionalities to be integrated into the tool and development of the site architecture
AC2	1	4 CWs (including 1 who also participated in AC1), 2 HSSPs (both of whom also participated in the AC1), 2 caregivers (including 1 who also participated in AC1)	Choice of functionalities that were not consensual
CoD6	3	4 CWs, 3 HSSPs, 3 caregivers	Functionalities and content development
CoD7	5	3 CWs, 2 HSSPs, 5 caregivers	Functionalities and content development
CoD8	3	4 CWs, 2 HSSPs, 7 caregivers	Functionalities, content development, and pretesting
AC3	2	4 CWs (including 1 who also participated in AC1), 2 HSSPs (both of whom also participated in the AC1), 2 caregivers (including 1 who also participated in AC1)	Exploration of the prototype, choice of realistic functionalities, and discussion on the content

[a]CoD: co-design sessions.

[b]The number of preparation sessions was not defined in advance but rather on an as-needed basis, depending on the evolution of the prototype and the complexity of the results analysis.

[c]AC: advisory committee.

[d]CW: community workers.

[e]HSSP: health and social service professional.

Data Collection

Several stages of the QADA project focused on exploring the ways in which co-design can support the development of inclusive eHealth tools that contribute to reducing SHIs:

1. Preparatory meetings for the CoDs (including the AC) by the research team (n=24). These provided information regarding the efforts made to ensure the optimal mobilization of participants, obtain consensual decision making, and choose the information to be presented. The resulting documents (CoD planning) and the audio recording of these meetings were used as raw data for analysis.
2. CoDs (n=8 CoDs and 3 working sessions of the AC). These sessions produced information about the co-design process. The sociodemographic data of the participants (provided by them) and the audio recordings of these meetings served as raw data for the analysis.
3. Co-design postsession debriefing meetings (n=11). These meetings helped to quickly identify researchers' perceptions

of the co-design process. Note taking during debriefing and audio recordings also served as raw data for analysis. These meetings took place immediately after each CoD.

Data Analysis

The analysis plan followed the method proposed by Miles and Huberman [44,45]. In this study, this resulted in a written summary of each document and audio recording from preparations of CoDs, the CoDs themselves, and debriefings. Deductive coding was performed to link the content relating to each conversion factor, including the co-design process, by using the MAXQDA software (Verbi) [46]. To refine the subthemes for the co-design process and prioritize the most relevant outcomes, the arguments for and against using a co-design approach in the development of an eHealth tool were transformed into analytical questions and used as a basis for presenting the results [47]. Textbox 1 presents these analytical questions. The same method was used to apply the democratic process guiding principles drawn from Bonvin and Farvaque [34] on the results (Textbox 2).

Textbox 1. Analytical questions based on arguments for and against using a co-design approach to develop an eHealth tool and the associated results.

- Can co-designers significantly contribute to the development of an eHealth tool?
  - Are they able to express their needs and preferences?
  - Are they able to contribute even when they lack technological skills?
  - Does their contribution lead to innovation?
  - Can co-developers provide key information with a potential rapidly improve the effectiveness of the eHealth tool?
- Does the co-design approach involve only a small number of participants, which consequently makes the technology developed overspecialized and relevant only for a few people?
- Does co-design provide a reciprocal relationship between the research team and the people involved in the project?
  - Does the FU's involvement increase the cultural and logistical relevance of the action in the community, support empowerment, and promote knowledge transfer?
  - Do people who engage in co-design processes perceive beneficial effects on their physical and psychological health, self-confidence, self-esteem, empowerment, and social relationships or, on the contrary, physical and psychological exhaustion, stress, and financial loss?
  - Do people experience disappointment with the results of the process?
- Do disagreements within the group give rise to negotiations and compromises that lead to inconclusive results?
- Is involving the FUs an easy task to accomplish for researchers and designers in terms of time, energy, and competence and is the added value of this type of approach clear and explicit?

Textbox 2. Analytical questions based on guidelines proposed by Bonvin and Farvaque for the development of an eHealth tool in a democratic way.

- Have people at risk of SHIs had space to express themselves?
- Has the decision-making power been shared with people at risk of SHIs?

Ethical Considerations

This project was approved by the Comité d'éthique de la recherche des Centres de santé et de services sociaux de la Vieille-Capitale (Research Ethics Committee of the Health and Social Service Centers of the Old Capital).

Results

Presentation of Results

The results are first presented in response to the arguments for and against using a co-design approach in the development of an eHealth tool. We then proceeded to lay out the manner in which we applied the democratic process, as described by Sen in our experiment. Finally, these results are discussed in relation to the research question, that is, in what way can co-design support the development of inclusive eHealth tools and consequently contribute to the reduction of SHIs. We begin by presenting the co-designer's sociodemographic characteristics.

Co-Designers' Characteristics

A total of 78 co-designers participated in 11 CoDs, 24 preparation sessions, and 11 debriefing sessions. Table 2 presents the characteristics of the people who contributed to this research.

Table 2. Description of co-designers (N=78).

Sociodemographic characteristics	Caregivers (n=30)	Community workers (n=26)	Health professionals (n=18)	Research team (n=4)
Sex, n (%)				
Female	26 (87)	20 (77)	18 (100)	4 (100)
Male	4 (13)	(23)	0 (0)	0 (0)
Age (years)				
Range	42-88	24-66	29-53	33-45
Mean (SD)	77.9 (11.0)	44.8 (12.3)	39.6 (7.9)	40.7 (5.4)
Education level, n (%)				
Elementary school	1 (3)	0 (0)	0 (0)	0 (0)
High school	10 (33)	1 (4)	0 (0)	0 (0)
College	4 (13)	4 (15)	6 (33)	0 (0)
Vocational studies	1 (3)	0 (0)	3 (17)	0 (0)
University	12 (40)	21 (81)	9 (50)	4 (100)
None	1 (3)	0 (0)	0 (0)	0 (0)
N/M[a]	1 (3)	0 (0)	0 (0)	0 (0)
Age of the person cared for (years)				
Range	61-96	N/A[b]	N/A	N/A
Mean (SD)	78.2 (9.9)	N/A	N/A	N/A
Relationship with the person cared for, n				
Children	8	N/A	N/A	N/A
Sibling	3	N/A	N/A	N/A
Spouse	17	N/A	N/A	N/A
Friend	2	N/A	N/A	N/A

[a]N/M: not mentioned by the co-designers.

[b]N/A: not applicable.

Can Co-Designers Contribute Significantly to the Development of an eHealth Tool?

The majority of co-designers made a significant contribution at some stage of the prototype development. One co-designer even jokingly requested copyright:

> *Can we copyright this? (laughing).* [Community worker CoD5]

Were They Able to Express Their Needs and Preferences?

All co-designers (caregivers, HSSPs, and community workers) except 2 caregivers identified the priority needs (19 needs), analyzed eHealth tools to identify the desirable and missing features, created functionalities to address the unmet needs (25 functionalities), and created the website's architecture (3 prototypes of low fidelity) and content (text written on the website, a questionnaire for the caregiver to identify their needs, a glossary to associate resource data, keywords that can potentially be used by caregivers in a search tool, and scenarios for potential videos). Some co-designers had a very clear vision of what the tool might look like:

> *A dropdown menu on the homepage. They will click on "issue affecting the older person" (Alzheimer's disease) and they'll find all the services available to the senior and the caregiver. They will recognize themselves based on the profile of the person to whom they provide care and through the search that they do for the senior. Menu based on the entry issue. There would be two entries, one "I am a caregiver" and the other "the person I provide care to has …. By diagnosis (for example, Alzheimer's disease) or by issue facing the older person (for example, forgets to take their medication). Both! Both possibilities.* [Community worker CoD2]

Were Co-Designers Who Were Lacking Technological Skills Able to Contribute?

Although the majority of co-designers contributed to the development of the prototype, 2 caregivers were unable to participate in the design activity. Despite the efforts of the research team members to clarify the objectives of the working session, these 2 individuals did not seem to be able to make the link between their experiences and the objectives of the session. The team members understood that what these caregivers needed

most was to talk and they lent an attentive ear to meet their needs:

> *What do you think of this website?* [Member of the research team (CoD3)]

> *At my house, the shower is downstairs and he (spouse) can't use it. I bought armrests for the toilet.* [Caregiver (watching an image of a bathroom with special needs equipment proposed by the website)]

> *Ms. X didn't know why she was there. KL repeated several times that she could leave and made sure that they had her informed consent. It was hard for her.* [Debriefing, CoD3]

In addition, it seemed difficult for some co-designers (both caregivers, HSSPs, and community workers) to think in numerical terms, especially at the beginning of the process, during brainstorming activities to create functionalities targeted to meeting needs (CoD1 to CoD4). On several occasions, the solutions or reflection focused on the health system and current services rather than on the use of digital technology to meet the caregiver's needs. The research team repeatedly refocused the co-designers on the objective of developing a digital tool targeted to meet the needs of caregivers:

> *The system should coordinate their services with those of community organizations, including references.* [Community worker, CoD1]

> *Each CLSC should have a form where individuals can indicate their needs and streamline efforts where they are most needed. They could adapt services to needs.* [Caregiver, CoD2]

> *Technology is limited. TV advertisement worked really well. In Saguenay, the newspaper works well.* [Community worker, CoD3]

> *Financial compensation for meeting with the caregiver and conducting an assessment.* [HSSP, CoD4]

> *A follow-up by telephone with caregivers at least once per year.* [Caregiver, CoD4]

> *Social workers also used to think in terms of social services and it took some reframing in order to think digital.* [Debriefing, CoD4]

However, all HSSPs, community workers, and 28 caregivers contributed in one way or another to the development of the prototype.

An important finding was that caregivers who were less comfortable with technology and less talkative because of their difficulty in contributing directly to the designing of the tool explicitly agreed to take on the role of guardians of the literacy level:

> *Caregiver (CoD6): I don't work hard. I'm a caregiver. I see how hard you work in organizations.*

> *Community worker: we work for you.*

> *Member of the research team: However, you play an important role in the group by saying "I don't*

understand this word" or "What does this mean?" Your contribution is essential in this respect.

As a result, this caregiver went from being silent to regularly participating in her subgroup:

> *When you say home transportation, you mean that someone comes to our home to replace us? (…) Support group: what do you mean? (…) What is GMF?* [Caregiver, CoD6]

This contribution of caregivers both in speaking out and in ensuring that the tool respects the eHealth literacy level of FUs was then reinvested in the following CoDs.

Did Co-Designers' Contribution Lead to Innovation?

The members of the AC raised the paradox of having, on one hand, an essential need to move beyond the usual reference frameworks to create an innovative eHealth tool and the challenge, on the other hand, of effectively moving beyond these reference frameworks in the search for solutions and creation activities, despite an awareness of this need:

> *We all come up with the same idea, i.e. a website offering information. We're having a hard time thinking outside the box. We stick to the tried and tested.* [Researcher, AC1]

> *We shouldn't repeat what we know already, it will make no difference.* [HSSP, AC1]

> *Websites, apps, video games, video clips, surveys, quizzes, writing letters. I personally am not able to think creatively.* [Researcher, AC1]

Were Co-Developers Able to Provide Key Information to Rapidly Increase the Effectiveness of eHealth Tool?

At different times and at different levels, the co-designers brought up key information that the team would not have thought to incorporate on its own. For example, the idea that the location of community resources must be presented by district, the importance of obtaining results with no more than 2 clicks, the fact that resources must be both self-regulated and solicited, and the fact that caregivers sometimes have to do research for several Quebec regions or for several profiles because they provide care to more than one person:

> *For Montreal, we're facing the challenge of postal codes, CSSS territory: most organizations use the old CSSS system, but some are neighbourhood-focused. Searching by distance can't work for Montreal because an organization may be near your home, but you may not have access to it because you're covered by one that is 20 km away. Distance doesn't mean anything.* [Community worker, CoD6]

> *I just want to browse fast, I don't have time. I don't have the time to sign up.* [Caregiver, CoD5]

> *Community worker, CoD6: Will the tool cover public and private service organizations?*

> *Member of the research team: people will sign up with user verification. Community stakeholder: your data bank will be half full. We must also solicit resources.*

The caregiver thought that she had to create a profile for St-Hyacinthe and for Drummondville because she provides care to two individuals. [Member of the research team, debriefing CoD8]

Does the Co-Design Approach Involve Only a Small Number of People, Which Consequently Makes the Technology Developed Overspecialized and Relevant Only for a Few People?

The co-designers, except for the research team, are potential FUs of the tool developed. The particularity of the methodology used (ie, having a different group of co-designers at each CoD) enabled the contribution of a greater number of co-designers to the project. As Table 1 shows, 74 of them were divided into 3 categories: caregivers, community workers, and HSSPs. Owing to their different roles, these 3 types of FUs had different experiences (and perspectives) in relation to caregivers' help-seeking process. The sociodemographic characteristics of the co-designers highlight a variety of educational levels (primary to university level) and age range (24-88 years). However, there is a predominance of women. It appears that the co-design approach can be used with the involvement of a greater number of participants but, more importantly, with a greater diversity of participants to limit the risk of creating an overspecialized eHealth tool. However, this method has a set of challenges. Indeed, the co-designers had to agree to build the tool based on previous decisions that were not their own. The research team had to take about 30 min at each CoD to explain the progress of the project and the decisions made by the previous groups. In addition, this method did not allow the co-designers to develop their co-creative skills over the long term and limits the possibility of developing a long-term relationship of trust.

Does Co-Design Provide a Reciprocal Relationship Between the Research Team and the People Involved in the Project?

Does the FU's Involvement Increase the Cultural and Logistical Relevance of the Action in the Community, Support Empowerment, and Promote Knowledge Transfer?

The research team and practitioners gained knowledge about the caregiver's help-seeking process and the reality and issues facing community workers and HSSPs. The latter also discovered several other digital tools on caregiving, took advantage of the expertise of caregivers in a context other than the supportive-based relationship, shared advice with each other, and discovered existing services in regions other than their own:

> *It's already quarter to three. Time flies. I like this, I learn things.* [HSSP, CoD5]

> *This is the first time I've participated in training as quick and well done as this. It's not a waste of time.* [Community worker, CoD1]

The participation of caregivers made it possible to adapt the tool's functionalities to their help-seeking process. This contribution was a unique learning experience for the research team, HSSPs, and community workers because caregivers described how they searched for help on the internet, especially in the physical and psychological state they find themselves in when they engage in help-seeking. This influenced, among other things, the development of the prototype in terms of functionality (as few features as possible so that the desired information would be found with no more than two clicks), the home page (search box highlighted on the home page), quick access to a contact person (phone number), the incentives for initiating contact with formal resources (eg, caregiver testimonial video and virtual tour of the site), and the importance of exchanges with other caregivers to initiate the help-seeking process:

> *The forum is of no help to me as a caregiver. Because if I post something, this makes one more page I have to check. I prefer talking with someone right away. I need contact, quickly. No time for the forum or the chat. Things are difficult today, I want to talk to a real person.* [Caregiver, CoD5]

> *I did a search on the X website. But sometimes it takes too long. We do it in the evening, we're already tired, it's too taxing. I think the project must ensure that we can get to the information rapidly and that we can take rapid action.* [Caregiver, CoD7]

Did the Participants in the Co-Design Process Perceive Any Beneficial Effects on Their Physical and Psychological Health, Self-Confidence, Self-Esteem, Empowerment, and Social Relationships or, on the Contrary, Physical and Psychological Exhaustion, Stress, and Financial Loss?

The research team set up activities that were intended to be effective and fun. They wanted to create a friendly environment where everyone was free to drink, eat, and move around as they pleased. The objective was to create the prototype in a serene and safe environment. The recordings reveal that co-designers laughed together on several occasions.

> *Be very comfortable to move around at any time. We want it to be relaxed.* [Research team, CoD6]

Caregivers, for their part, discovered new services that exist in their region and shared their experiences with other caregivers, and some felt they were contributing for the benefit of caregivers in the future:

> *I think that this is promising for future caregivers. It's a way to give back. Yes, they'll be better informed, more knowledgeable, and more thanks to the project. It will be easier for them and better for the person receiving care.* [Caregiver, CoD2]

On the other hand, 4 caregivers and 1 HSSP verbally expressed emotions of anger, sadness, guilt, and fatigue unrelated to their participation in the project but related to their work or role. The team adjusted the schedule to allow for time to express these emotions. In addition, the team ensured that each time, the person had the necessary support following the CoD (often with the community worker or HSSP who was already on the site):

> *Resources are lacking. (…) There's a lack of personnel; I wasn't able to find anyone to cover*

nights. I slept 3-4 hours per night (crying). I slept in the basement. Human resources, self-employed workers, are lacking. We have to take this into consideration. Even with the best tools, if there is no one to provide the service, we're running into a wall. The population is aging. The money in the healthcare system is poorly distributed. [Caregiver, CoD7]

Do People Experience Disappointment With the Results of the Process?

At the time of writing, the co-designers had not yet seen the final result, with the exception of the members of the AC. The latter did not explicitly express satisfaction or disappointment with the prototype.

Did Disagreements Within the Group Give Rise to Negotiations and Compromises That Led to Inconclusive Results?

Three major themes were the subject of dissension among the co-designers. The first concerns the difference between the needs identified in CoD1 and CoD2. The AC1 working session made it possible to decide to keep all the needs identified in the 2 CoDs:

What is the point in eliminating needs? Do we have to have a limit? [Researcher (AC1)]

We can miss out on something. [HSSP]

If we get rid of some needs and it turns out that the tool doesn't meet a given need, maybe it would have met another. We don't know. We might eliminate needs that are easy to respond to. [Researcher]

I would keep them all and some can probably be grouped together. [Community worker]

The second concerns the desire of caregivers to have a space for discussion with other caregivers, whereas community workers and HSSPs feared that this would open the door to people who are ill-intentioned and who would take advantage of the caregiver's moment of vulnerability or offer the caregiver misguided advice. This discussion started in CoD3. This dissension was discussed at AC2 without finding a solution. However, the committee mandated the research team to find a solution that was acceptable to all. The dilemma continued until CoD7, where the co-designers proposed a solution that met the AC's expectations, that is, to give organizations the opportunity to indicate on the website that they do twinning between caregivers (the twinning is therefore supervised by a professional).

We have to meet with people before they get too tired; right now, contacts are made while people are exhausted. If platforms for discussion are available, initial contact can be made and when things aren't going so well, I know already that I have things in common with someone. (...) [Caregiver (CoD7)]

I think matchmaking should be organized in regions where the need is felt and that local organizations should support it. This could be a service offered by these organizations. In their service offer they should put "matchmaking service to encourage twinning

with a caregiver" and there, the caregiver can contact the organization to use the service. [Community worker]

The third dissension concerned offering a system, on the website, for evaluating the services offered by organizations (similar to hotel and restaurant ratings). Organizations mentioned that such reviews are subjective and could affect their funding. Caregivers also pointed out that the work environment in their home is not always conducive to doing a good job and that it was not necessarily the worker's fault. There is also a major difference between the service and the person providing the service. The caregiver may not be happy with a person who nevertheless provided high-quality service. It was a community worker who found a consensus solution:

It's true that we don't always dare to use services. If I leave and someone else is to provide care to the person that I love more than anything in the world, this caregiver enters my privacy, and this is disturbing. Do they have a criminal record, all employees (should be) screened? Providing elements of trust to caregivers. [D7]

As a result, a number of statements that organizations can check if they apply to them were included. They appear in the caregiver's search results. For example, all employees and volunteers have received training on neurocognitive disorders or the criminal record of all employees and volunteers is checked.

Is Involving the FUs an Easy Task to Accomplish for Researchers and Designers in Terms of Time, Energy, and Competence, and Is the Added Value of This Type of Approach Clear and Explicit?

This question will be addressed in the *Discussion* section.

Co-Design as a Democratic but Imperfect Process

Have People at Risk of SHIs Had the Space to Express Themselves?

The team sought to ensure that speaking time was shared during the CoDs, in particular by involving a number of facilitators. In addition, at the beginning of each session, clear instructions were issued to encourage equality of expression:

Give everyone a chance to speak; those who tend to take a lot of space are asked to give others a chance, and those who are reluctant to speak are invited to make an effort to participate. We would like to hear what each of you has to say. [Team member, CoD1]

In addition, the facilitator tried to reach out to less-talkative participants:

I see you nodding (to a person who didn't speak much). [Team member, CoD2]

However, it was also important to respect people's personalities:

I asked people who don't speak up if they have something to add in response to what was just said. At the same time, it's putting them on the spot. Not

XSL•FO

RenderX

easy. We want people to talk, but we don't want to force them. [Debriefing, CoD4]

Other measures were put in place to support discussion, including limiting CoD groups to 6 to 12 people and introducing subgroup activities with 2 to 3 participants and a facilitator:

In small groups, everyone spoke, unlike in the plenary session where one lady did not speak at all. People liked this. [Debriefing, CoD3]

In particular, the team anticipated that community workers and HSSPs would take up a lot of space, unlike caregivers. Initially, the facilitation plans called for the use of small groups with a facilitator and dividing participants according to their roles (eg, community workers together, HSSPs together, and caregivers together). In fact, the team found an equivalent level of input in the first CoD. However, the cross-fertilization of expertise within a subgroup also appeared to generate more creativity:

For the next co-design session, I'd mix up participants. We have to foster creativity and for this reason, it's good to mix participants to stimulate the flow of energy. We want divergent thinking. [Preparation, CoD4]

In addition, in mixed subgroups, some HSSPs and community workers sometimes helped caregivers to speak out:

In the definition, the social worker got the caregivers to speak. She asked them "What made you realize that you were caregivers?" They brought up important notions, key elements. [Debriefing CoD6]

However, this was not done always:

Team member to a caregiver CoD6: Does this sound familiar to you?

Caregiver: yes... (tried to answer, but was interrupted by an HSSP).

The team's observation was that speaking out depended on personalities more than the role or status of each person (a caregiver, an HSSP, or a community worker), and it was therefore difficult to plan in advance:

Are we going to fail to get the caregivers' opinion because healthcare and community workers will be speaking at the time? We can't anticipate this. [Preparation, CoD2]

In short, the objective of the workshop, the need to combine expertise or absence thereof, and the personality of co-designers were all factors that influenced the relevance of using mixed or homogeneous subgroups. For this experiment, the latter failed to guarantee shared speaking time. The presence of a facilitator in small groups was more effective in ensuring equal opportunity for expression.

Was the Decision-Making Power Shared With People at Risk of SHIs: A Shift From Ideal to Pragmatic Considerations

Initially, the team proceeded with a concern for making participants the main decision makers in the development of the tool. It aimed to be as neutral as possible so as not to influence participants:

Caregiver CoD5: It depends on your needs (speaking of the website).

Team member: We don't have specific needs. It's up to you to decide.

(Speaking of co-designers) They are the ones creating it (the tool). [Preparation, CoD2]

However, it must be admitted that several barriers led to the team becoming increasingly involved in decision making. First, the team indirectly influenced the CoDs by preparing facilitation and providing materials to support the participation of the co-designers:

We'll start with a list of needs drawn from the literature and the pilot project. But we need to reformulate these needs in order to facilitate group discussion. Ideally, the reformulation would be done by the group but because we're short on time, the research team will do this (clean up) and provide the group with a list. Therefore, the research team necessarily filters information. However, the group will have the possibility to take out, add, or reformulate things. [Preparation, CoD1]

Second, by condensing data from the sessions required to analyze the data to plan the next meeting:

The next step is juxtaposing results. Each subgroup will transcribe this, and we'll look at this at the preparatory meeting Thursday morning. [Debriefing, CoD6]

Analyzing the results is another factor. Although the research team had the rigor to cross-reference the analyses, this cross-reference was carried out by the research team:

Everyone will have to review X's analysis to confirm that it corresponds to what we've addressed. [Preparation, CoD4]

The AC's initial role was to be the final decision maker in cases where the groups had divergent opinions or if proposals appeared unrealistic. It also had the responsibility to ensure that the decisions made by the co-designers in the development of the prototype were respected. However, the AC was not an operational committee and, in this sense, it took place only 3 times in the process:

The advisory committee meeting is far, choices will probably have to be made and we're the ones who will make them in the end... (to prepare the next co-design session). [Preparation, CoD4]

Moreover, there can be divergent opinions within the AC itself. Among other things, caregivers wanted to have the possibility to be matched with other caregivers; however, HSSPs and community workers feared that this would expose them to ill-intentioned people. This difference of opinion was brought to the AC, which was incapable of deciding. The research team proposed changing the weighting of opinions in favor of the caregivers. However, the AC mandated the research team to mediate a solution. The team brought the debate back to the CoDs twice and arrived at a solution.

Finally, the co-design approach requires a substantial amount of time. Time constraints (the project duration and the financial support to assume the number of CoDs) led the team to make decisions in view of subsequent sessions and move the prototype forward. This constraint was anticipated from the outset:

> *From a practical point of view, decisions regarding data will probably be taken by the research team. Although we'd like all decisions to be made in the co-design context, we must recognize that time (and the fast pace of the project) will force us to make decisions outside of co-design.* [Preparation, CoD1]
>
> *Team member 1: Yes, we are pressed for time right now. I don't feel comfortable. We're making plenty of decisions. [Preparation CoD7]*
>
> *Team member 2:You're right, we're moving away from co-design.*
>
> *Team member 3: We want to push because there is a deliverable. I'm getting more and more uncomfortable. We will have no choice but to make a lot of decisions. [Discussion between team members, Preparation CoD7]*

Once the team realized that, from a pragmatic standpoint, not everything could be done within the scope of CoDs, it had to set apart topics that had to be discussed in sessions and those that were less impactful on the overall quality of the co-design process, which it could decide on itself.

Issues pertaining to the methodology were left to the research team. Some features deemed unrealistic in this study were also removed by team members. The importance of the final decision and the availability of information were also the factors that determined whether or not to discuss an issue in a CoD:

> *Team member 1: We're in the process of making an important decision. [Preparation CoD7]*
>
> *Team member 2: This aspect should be addressed in a co-design setting. This is important.*
>
> *Team member 1: Being kept abreast of activities in my region…We wouldn't have gone any further in our thinking. [Preparation CoD7]*
>
> *Team member 2: Shall we take this to co-designing? (…)*
>
> *Team member 1: We have no choice since we don't know (…) We've never discussed this from this vantage point. To be moved to the co-design setting. [Preparation CoD7]*

Discussion

Principal Findings

In this section, we will answer the research question in light of the aforementioned results, that is, how co-design can support the development of an inclusive eHealth tool for caregivers of functionally dependent older persons and consequently contribute to the reduction of SHIs.

The first question: Can people in situations of vulnerability (caregivers in this case) with no expertise in design contribute in a significant manner to the design of an eHealth tool and provide insight to help make the tool more inclusive?

In the Introduction section, we presented the 2 dominant schools of thought on this question. One posits that people can contribute meaningfully by expressing their interests, attitudes, beliefs, values, and expectations, by demonstrating their capacities [10,18], and by describing the context in which they will use the eHealth tool [25]. In addition, they can identify what is useful and relevant [21]. Their participation may also lead to innovation [24]. On the other hand, skeptics of the co-design approach timidly argue that people may not be aware of their needs, be able to express their needs, or want to discuss their needs in a group setting [28]; that the technology developed can be overspecialized and relevant only to a few [28]; and that people will have difficulty contributing significantly if they do not have technological skills [15]. Our position? Somewhere between these 2 paradigms, we advocate the use of co-design while keeping a critical eye. On the basis of our experiment, some caregivers did have skills and knowledge in digital technology, which accounts for the important role they played in the development of the prototype. Two caregivers were unable to contribute to the development of the prototype in any way. However, with the exception of these 2 individuals, all caregivers (including HSSPs and community workers) made a significant contribution to the process. Some reframing by the research team was necessary to ensure that the solutions found were linked to the digital tool rather than focused on the current health system and services. Nevertheless, most co-designers made a key contribution, some directly on the prototype by offering creative ideas and key information and others indirectly by ensuring consistency between the tool and an adequate level of literacy and by respecting the caregivers' help-seeking process. If the use of a co-design approach, particularly with people at risk of SHIs (in this case caregivers), facilitates the integration of other conversion factors such as eHealth literacy and help-seeking, it therefore contributes to the development of universally accessible tools and potentially contributes to curbing SHIs. The co-design approach maximized this participation beyond what could have been expected by the research team, that is, key information was provided that could not have been anticipated by interview questions, for example. This tends to support the view that involving FUs in the development of a prototype allows for going beyond what can be seen in an observation or heard in an interview [16,17]. In this sense, the time and energy invested in the co-design yield added value.

However, let us take this reflection a step further and answer the following question: Can co-design be considered a democratic process, as described by Sen? Can an equal distribution of decision-making power be implemented between the different co-designers? In all humility, the answer to this question is more nuanced.

In this experiment, the role of the research team in co-designing and its neutrality in the research project were ambiguous. Its original intent was to remain neutral to leave the decision-making power entirely in the hands of the co-designers; however, its position evolved to where the team defined itself as a co-designer in its own right and, finally, to become a more

important decision maker, all the while respecting the decisions made by the other co-designers. The underlying reflection varied over time and between researchers. It seemed impossible for the research team to not intervene directly in the development of the tool and therefore to make decisions without debate with the co-designers. What practical impact did this choice have on the democratic process, as understood by Sen?

Bonvin and Farvaque [34] invited us to question our interpretation of what limits or hinders democratic development and, consequently, of what can be an obstruction factor in the development of capabilities. They address the importance of people having access to debate, facilitating their expression, and including their views in the decision-making process. In this experiment, caregivers who wanted to contribute to the development of the eHealth tool were all welcomed regardless of their level of education, digital literacy, or technological skills. The research team put in place various means to ensure free discussion and the expression of different points of view. Participants' values, including those related to the help-seeking process, were incorporated into the tool. Two decision-making power—in other words, democratic—issues (obstruction factors) arose. The first was a divergence of choices between caregivers and stakeholders (community and health network). It was decided to seek a solution acceptable to all parties through mediation. These tensions can be seen as democratic limitations resulting from the choice to involve stakeholders in the design process, but they can also be seen as creative resources that are part of the co-design method because they bring to light issues that would have emerged sooner or later [48].

The second was when the time allocated for the project ran out. At that point, some issues had to be left out of CoDs. The lack of time prompted the team to discern what was essential to designing an inclusive tool versus decisions that did not have a democratic stake. It would have been valuable (and even important) at that time if a caregiver could contribute to this reflection. However, the presence of a caregiver to help with the preparation meetings would have been impossible because of availability constraints. The team had 1 to 4 preparation meetings per CoD, often in less than a month. We anticipated the negative effects discussed by Attree et al [27]. It would be useful to single out some key issues to guide researchers on what should be addressed in co-design discussions and what can be left out to make optimal use of this approach in a real-world context. For example, is this decision consistent with all the information we have received so far in the context of the co-design process? Does this decision benefit some people to the detriment of others?

Looking at the spectrum of public participation in the International Association for Public Participation [49], it can be broadly said that, on the whole, stakeholders' involvement was limited to having a collaborative role, that is, the research team sought the advice and innovative ideas of the FUs and, to the extent possible, took them into account in the decisions it made. Could it have been otherwise? Theoretically, a genuine sharing of power requires participants to have an enabling role. From a pragmatic standpoint, is this one of the limits of co-design, and even an epistemological limit?

Individual interviews with co-designers would complement this reflection. How did they experience the co-design process? How did they feel about the project? Did they feel they had the opportunity to express themselves freely? This is the subject of a doctoral thesis by one of the members of the research team (MT). Data analysis is currently underway.

Strengths and Limitations

To ensure the credibility of this study, the data collection was spread over a period of one year and involved a variety of participants (caregivers, community workers, HSSPs, and coresearchers) having various profiles (age, comfort level with technology, etc). Triangulation of the data was done through the use of various sources of information, including recordings of the preparation sessions, co-design, debriefing sessions, artifacts produced, and notes taken during the work sessions. In addition, the accuracy of the summary documents was verified by a member of the research team (an anthropologist and research professional) who participated in the working sessions, co-designing, and debriefing. She checked the accuracy of 10% of the documents at random. This study is the subject of a thesis and is therefore supervised by a thesis committee composed of 4 university researchers in the fields of education and health. However, it also has limitations. Although we sought to include a variety of caregiver profiles, most co-designers were women and White. This may have had an effect on the interpretation of results concerning, among other things, the process of help-seeking, which is a cultural construct [50]. In addition, the majority of participating caregivers were retired and were already service users. Caregivers among the active population or caregivers at the beginning of the help-seeking process could have made different choices. Moreover, 12 of the 30 caregivers who participated had a university degree. According to Sen, only individual preferences that pass the democratic debate test deserve to be supported by public action insofar as this debate leaves as much space as possible for the individuals concerned. Although all caregivers are equally concerned about the tool as the others, in the context of fighting SHIs, the higher education of some may have influenced certain choices, including the content developed.

Conclusions

The use of a co-design approach does not guarantee that participants at risk of SHIs will bring innovation, but it does guarantee, if their participation is promoted, that the tool developed will respect their process of help-seeking and their literacy level and potentially contribute to curbing the existing SHIs. Once the eHealth tool developed as part of this project is finalized, a usability study comparing people at risk of SHIs and people who are not at risk will confirm or invalidate this hypothesis. However, this approach requires time, and it is difficult to achieve the time and money constraints associated with research grants. Key issues supporting the reasoning of researchers and designers on what must be absolutely debated in the development of an inclusive eHealth tool could allow for optimal and realistic use of time.

Conflicts of Interest

None declared.

References

1. eHealth. World Health Organization. 2004. URL: https://www.who.int/healthacademy/media/en/eHealth_EB-en.pdf?ua=1 [accessed 2020-02-24]
2. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010 Dec 17;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]
3. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
4. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
5. Health Information Technology Initiative: Major Accomplishments 2004-2006. US Department of Health and Human Services. 2007. URL: http://www.providersedge.com/ehdocs/ehr_articles/ Health_Information_Technology_Initiative-Major_Accomplishments_2004-2006.pdf [accessed 2020-10-26]
6. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows ehealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
7. Latulippe K, Hamel C, Giroux D. Social health inequalities and ehealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
8. Hyppolite S. Comprendre Et Agir Autrement Pour Viser L'équité en Santé Dans La Région De La Capitale-nationale, Rapport Du Directeur Régional De Santé Publique Sur Les inégalités Sociales De Santé. Centre Intégré Universitaire De Santé Et De Services Sociaux De La Capitale-nationale. 2012. URL: http://www.ciusss-capitalenationale.gouv.qc.ca/sites/ default/files/rapportiss_versionintegrale.pdf [accessed 2020-02-24]
9. Bacigalupe G, Askari SF. E-Health innovations, collaboration, and healthcare disparities: developing criteria for culturally competent evaluation. Fam Syst Health 2013 Sep;31(3):248-263. [doi: 10.1037/a0033386] [Medline: 24059273]
10. Baur C. An analysis of factors underlying e-health disparities. Camb Q Healthc Ethics 2008;17(4):417-428. [doi: 10.1017/S0963180108080547] [Medline: 18724881]
11. Benton L, Johnson H. Widening participation in technology design: a review of the involvement of children with special educational needs and disabilities. Int J Child-Comput Interact 2015 Jan;3-4:23-40. [doi: 10.1016/j.ijcci.2015.07.001]
12. Dell'Era C, Landoni P. Living lab: a methodology between user-centred design and participatory design. Creativity Innov Manag 2014 Mar 18;23(2):137-154. [doi: 10.1111/caim.12061]
13. Jagosh J, Pluye P, Macaulay AC, Salsberg J, Henderson J, Sirett E, et al. Assessing the outcomes of participatory research: protocol for identifying, selecting, appraising and synthesizing the literature for realist review. Implement Sci 2011 Mar 20;6:24 [FREE Full text] [doi: 10.1186/1748-5908-6-24] [Medline: 21418604]
14. Goodyear-Smith F, Jackson C, Greenhalgh T. Co-design and implementation research: challenges and solutions for ethics committees. BMC Med Ethics 2015 Nov 16;16:78 [FREE Full text] [doi: 10.1186/s12910-015-0072-2] [Medline: 26573410]
15. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
16. Steen M, Manschot M, de Koning N. Benefits of co-design in service design projects. Int J Design 2011;5(2):53-60 [FREE Full text]
17. Sanders EB. From user-centered to participatory design approaches. In: Frascara J, editor. Design and the Social Sciences. New York, USA: CRC Press; 2002:18-25.
18. Expanding the Reach and Impact of Consumer E-health Tools. Office of Disease Prevention and Health Promotion. 2006. URL: https://www.unapcict.org/resources/ictd-infobank/expanding-reach-and-impact-consumer-e-health-tools [accessed 2020-10-29]

XSL•FO

RenderX

19. Jagosh J, Macaulay AC, Pluye P, Salsberg J, Bush PL, Henderson J, et al. Uncovering the benefits of participatory research: implications of a realist review for health research and practice. Milbank Q 2012 Jun;90(2):311-346 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00665.x] [Medline: 22709390]

20. Ginossar T, Nelson S. Reducing the health and digital divides: a model for using community-based participatory research approach to e-health interventions in low-income Hispanic communities. J Comput-Mediated Commun 2010;15(4):530-551. [doi: 10.1111/j.1083-6101.2009.01513.x]

21. Patient Engagement. Canadian Institutes of Health Research. 2019. URL: http://www.cihr-irsc.gc.ca/e/45851.html [accessed 2020-02-24]

22. Selby JV, Beal AC, Frank L. The patient-centered outcomes research institute (PCORI) national priorities for research and initial research agenda. J Am Med Assoc 2012 Apr 18;307(15):1583-1584. [doi: 10.1001/jama.2012.500] [Medline: 22511682]

23. Burns KK, Bellows M, Eigenseher C, Gallivan J. 'Practical' resources to support patient and family engagement in healthcare decisions: a scoping review. BMC Health Serv Res 2014 Apr 15;14:175 [FREE Full text] [doi: 10.1186/1472-6963-14-175] [Medline: 24735787]

24. Bowen S, McSeveny K, Lockley E, Wolstenholme D, Cobb M, Dearden A. How was it for you? Experiences of participatory design in the UK health service. CoDesign 2013 Dec;9(4):230-246. [doi: 10.1080/15710882.2013.846384]

25. Groeneveld BS, Boess SU, Freudenthal A. Community-based co-design for informal care: bridging the gap between technology and context. IFAC Proceedings Volumes 2013;46(15):266-273. [doi: 10.3182/20130811-5-us-2037.00043]

26. Cyril S, Smith BJ, Possamai-Inesedy A, Renzaho AM. Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review. Glob Health Action 2015;8:29842 [FREE Full text] [doi: 10.3402/gha.v8.29842] [Medline: 26689460]

27. Attree P, French B, Milton B, Povall S, Whitehead M, Popay J. The experience of community engagement for individuals: a rapid review of evidence. Health Soc Care Community 2011 May;19(3):250-260. [doi: 10.1111/j.1365-2524.2010.00976.x] [Medline: 21138495]

28. Steen M. Tensions in human-centred design. CoDesign 2011 Mar;7(1):45-60. [doi: 10.1080/15710882.2011.563314]

29. Anderson LM, Adeney KL, Shinn C, Safranek S, Buckner-Brown J, Krause LK. Community coalition-driven interventions to reduce health disparities among racial and ethnic minority populations. Cochrane Database Syst Rev 2015 Jun 15(6):CD009905. [doi: 10.1002/14651858.CD009905.pub2] [Medline: 26075988]

30. Evans D, Pilkington P, McEachran M. Rhetoric or reality? A systematic review of the impact of participatory approaches by UK public health units on health and social outcomes. J Public Health (Oxf) 2010 Sep;32(3):418-426. [doi: 10.1093/pubmed/fdq014] [Medline: 20194176]

31. O'Mara-Eves A, Brunton G, McDaid D, Oliver S, Kavanagh J, Jamal F, et al. Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis. Public Health Res 2013;1(4):-. [Medline: 25642563]

32. Sen A. Repenser l'inégalité. Paris: Éditions du Seuil; 2000.

33. Picard F, Pilote A, Turcotte M, Goastellec G, Olympio N. Opérationnaliser La Théorie De La Justice Sociale D'amartya Sen Au Champ De L'orientation Scolaire : Les Apports D'une Étude Multicas Qualitative Et Comparative. Mesure et Evaluation en Education 2016 May 13;37(3):5-37. [doi: 10.7202/1036326ar]

34. Bonvin J, Farvaque N. Amartya Sen: Une Politique De La Liberté. Paris: Michalon; 2008.

35. Chiappero-Martinetti E, Venkatapuram S. The capability approach: a framework for population studies. African Population Stud 2014 Sep 2;28(2):708-720. [doi: 10.11564/28-2-604]

36. Latulippe K, Hamel C, Giroux D. Integration of conversion factors for the development of an inclusive ehealth tool with caregivers of functionally dependent older persons: social justice design. JMIR Hum Factors 2020 Aug 26;7(3):e18120 [FREE Full text] [doi: 10.2196/18120] [Medline: 32845242]

37. Latulippe K, Guay M, Éthier S, Sévigny A, Dubé V, Provencher V, et al. Supporting the process of help-seeking by caregivers of functionally dependent older persons through electronic health: protocol for a multicenter co-design. JMIR Res Protoc 2019 Apr 26;8(4):e11634 [FREE Full text] [doi: 10.2196/11634] [Medline: 31025956]

38. Creswell J. A Concise Introduction to Mixed Methods Research. Los Angeles, USA: SAGE publications; 2015.

39. Marston C, Renedo A, McGowan CR, Portela A. Effects of community participation on improving uptake of skilled care for maternal and newborn health: a systematic review. PLoS One 2013;8(2):e55012 [FREE Full text] [doi: 10.1371/journal.pone.0055012] [Medline: 23390509]

40. Bucki B, Spitz E, Etienne A, le Bihan E, Baumann M. Health capability of family caregivers: how different factors interrelate and their respective contributions using a Bayesian approach. BMC Public Health 2016 Apr 28;16:364 [FREE Full text] [doi: 10.1186/s12889-016-3027-8] [Medline: 27125282]

41. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [doi: 10.1037/0882-7974.18.2.250] [Medline: 12825775]

42. Vitaliano PP, Zhang J, Scanlan JM. Is caregiving hazardous to one's physical health? A meta-analysis. Psychol Bull 2003 Nov;129(6):946-972. [doi: 10.1037/0033-2909.129.6.946] [Medline: 14599289]

43. Yin R. Qualitative Research From Start to Finish. New York, USA: Guilford Publications; 2016.

44. Miles M, Huberman A, Saldana J. Qualitative Data Analysis. London, UK: Sage publications; 2013.

45.   Miles M, Huberman A. Analyse De Données Qualitatives. Brussels: De Boeck Université; 2003.
46.   MAXQDA. URL: https://www.maxqda.com/ [accessed 2020-02-04]
47.   Paillé P, Mucchielli A. L'analyse Qualitative en Sciences Humaines Et Sociales. Paris: A Colin; 2012.
48.   Grosjean S. Apports Et Limites Du Co-design Dans Le Développement D'innovation en Santé. Laboratoire D'innovations
      par et Pour les Aînés (LIPPA). 2019. URL: http://lippa.recherche.usherbrooke.ca/2019/12/19/
      apports-et-limites-du-codesign-en-sante/[Accessed [accessed 2020-10-01]
49.   Spectrum - Degré D'engagement du Public. International Association for Public Participation. 2018. URL: https://www.
      iap2canada.ca/resources/FR/Documents/AIP2Canada-Spectrum-FINAL-2016.pdf[Accessed [accessed 2020-10-01]
50.   Mojaverian T, Hashimoto T, Kim HS. Cultural differences in professional help seeking: a comparison of Japan and the US.
      Front Psychol 2012;3:615 [FREE Full text] [doi: 10.3389/fpsyg.2012.00615] [Medline: 23426857]

Abbreviations

AC: advisory committee
CoD: co-design session
FU: future user
HSSP: health and social service professional
QADA: Québec Ami des Aînés
SHI: social health inequality

Original Paper

Telemedicine Usage Among Urologists During the COVID-19 Pandemic: Cross-Sectional Study

Justin M Dubin[1], MD; W Austin Wyant[1], MSc; Navin C Balaji[1], BE; William LK Ong[2], MD; Reda H Kettache[3], MD; Malik Haffaf[4], MD; Skander Zouari[5], MD; Diego Santillan[6], MD; Ana Maria Autrán Gómez[7,8], MD, PhD; Hossein Sadeghi-Nejad[9,10], MD; Stacy Loeb[11], MD, MSc; James F Borin[12], MD; Juan Gomez Rivas[13,14], MD, PhD; Jeremy Grummet[15], MD; Ranjith Ramasamy[1], MD; Jeremy Y C Teoh[16], MD

[1]Department of Urology, University of Miami Miller School of Medicine, Miami, FL, United States

[2]Department of Urology, Penang General Hospital, Penang, Malaysia

[3]Department of Urology, Bachir Bennacer - Biskra Hospital, Biskra, Algeria

[4]Department of Urology, EHU 1er Novembre, Oran, Algeria

[5]Urology Department, Charles Nicolle Hospital, Tunis, Tunisia

[6]Department of Urology, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

[7]Department of Urology, Hospital Universitario Fundación Jiménez Diaz, Madrid, Spain

[8]Research Office of Confederacion Americana de Urologia, Buenos Aires, Argentina

[9]Department of Urology, Rutgers New Jersey Medical School, Newark, NJ, United States

[10]Department of Urology, Hackensack Meridian Health, Hackensack, NJ, United States

[11]Department of Urology and Population Health, New York University and Manhattan Veterans Affairs Medical Center, New York City, NY, United States

[12]Department of Urology, New York University, New York City, NY, United States

[13]Department of Urology, La Paz University Hospital, Madrid, Spain

[14]Department of Urology, Autonomous University of Madrid, Madrid, Spain

[15]Department of Surgery, Central Clinical School, Monash University, Melbourne, Australia

[16]SH Ho Urology Centre, Prince of Wales Hospital, Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China

Corresponding Author:
Justin M Dubin, MD
Department of Urology
University of Miami Miller School of Medicine
1120 NW 14th Street
Miami, FL, 33136
United States
Phone: 1 973 634 8677
Email: jdubin50@gmail.com

Abstract

Background: Prior to the COVID-19 pandemic, urology was one of the specialties with the lowest rates of telemedicine and videoconferencing use. Common barriers to the implementation of telemedicine included a lack of technological literacy, concerns with reimbursement, and resistance to changes in the workplace. In response to the COVID-19 pandemic declared in March 2020, the delivery of urological services globally has quickly shifted to telemedicine to account for the mass clinical, procedural, and operative cancellations, inadequate personal protective equipment, and shortage of personnel.

Objective: The aim of this study was to investigate current telemedicine usage by urologists, urologists' perceptions on the necessity of in-person clinic appointments, the usability of telemedicine, and the current barriers to its implementation.

Methods: We conducted a global, cross-sectional, web-based survey to investigate the use of telemedicine before and after the COVID-19 pandemic. Urologists' perceived usability of telemedicine was assessed using a modified Delphi approach to create questions based on a modified version of the validated Telehealth Usability Questionnaire (TUQ). For the purposes of this study, telemedicine was defined as video calls only.

Results: A total of 620 urologists from 58 different countries and 6 continents participated in the survey. Prior to COVID-19, 15.8% (n=98) of urologists surveyed were using telemedicine in their clinical practices; during the pandemic, that proportion increased to 46.1% (n=283). Of the urologists without telemedicine experience, interest in telemedicine usage increased from 43.7% (n=139) to 80.8% (n=257) during the COVID-19 pandemic. Among urologists that used telemedicine during the pandemic, 80.9% (n=244) were interested in continuing to use it in their practice. The three most commonly used platforms were Zoom, Doxy.me, and Epic, and the top three barriers to implementing telemedicine were patients' lack of technological comprehension, patients' lack of access to the required technology, and reimbursement concerns.

Conclusions: This is the first study to quantify the use, usability, and pervading interest in telemedicine among urologists during the COVID-19 pandemic. In the face of this pandemic, urologists' usage of telemedicine nearly tripled, demonstrating their ability to adopt and adapt telemedicine into their practices, but barriers involving the technology itself are still preventing many from utilizing it despite increasing interest.

Introduction

According to the American Telemedicine Association, telemedicine is the "use of medical information exchanged from one site to another via electronic communications to improve a patient's clinical health status" [1]. Telehealth, in comparison, is a combination of both telemedicine and remote nonclinical services such as provider training, administrative meetings, and continuing medical education. A 2016 survey by the American Medical Association [2] found that although most medical specialties use some form of telemedicine, videoconferencing is most used by physicians in emergency medicine, psychiatry, and pathology. Urology was found to be one of the specialties with the lowest rates of telemedicine and videoconferencing use. In a recent review, urologists commonly cited a lack of technological literacy, concerns with reimbursement, and resistance to changes in the workplace as barriers to the implementation of telemedicine in their practices [3].

In response to the COVID-19 pandemic declared in March 2020, the delivery of urological services globally has quickly shifted to account for the unprecedented mass clinical, procedural, and operative delays/cancellations, inadequate personal protective equipment, and shortage of personnel [4,5]. To accommodate the limited access to operative rooms, the surgical and urological communities took action: the American College of Surgeons recommended against elective surgeries and a multidepartmental group of urologists published urological surgery triage guidelines [6,7]. Globally, health care systems aggressively pushed for increased usage of telemedicine, with institutions such as New York University increasing their telemedicine use by over 4000% during this time [8]. The increase in telemedicine has inspired several organizations, including the European Association of Urology, to create recommendations for its professional usage, and some subspecialties have even provided guidelines for treating specific conditions via telemedicine [9,10]. The major shift in health care brought on by the COVID-19 pandemic has also seemingly incentivized all health care practices to more readily adopt telemedicine in their practices. The pandemic has exacerbated health inequities on a global scale, and we wanted to investigate whether urologists

and their practices were able to adapt to and overcome the previous barriers preventing them from prior telemedicine use. To our knowledge, however, the actual increase in urologists' usage of telemedicine during the pandemic has not been quantified. We hypothesized that, in response to the pandemic, urologists' use of telemedicine has increased. Our objective was to examine how urologists perceive the necessity of in-person clinic appointments, the usability of telemedicine, and the current barriers to its implementation.

Methods

We conducted a global, cross-sectional, web-based survey to investigate the use of telemedicine before and after the COVID-19 pandemic. Urologists' perceived usability of telemedicine was assessed using a modified Delphi approach to create questions based on a modified version of the validated Telehealth Usability Questionnaire (TUQ) [11]. The TUQ is a validated survey that was designed to assess both clinician and patient usability of telehealth implementation and services and previous studies have utilized it to evaluate telemedicine from the perspective of both the patient and the provider [12,13]. For the purposes of this study, telemedicine was defined as video calls only.

From April 15 to May 9, 2020, we distributed a 41-item survey to practicing urologists by email, via individual institutions, professional urology organizations, and Twitter. Distribution of the survey included an institutional review board (IRB)–approved introduction and invitation to complete the survey with a link that sent the user to the Qualtrics website to complete our survey. The survey was distributed via email and on Twitter. The survey included questions on urologist demographics, experiences with telemedicine prior to COVID-19, experiences with telemedicine during COVID-19, interest in telemedicine usage, barriers to telemedicine usage, and the TUQ to assess the telemedicine platform the provider was using. The survey was completely anonymous and took no longer than 8 minutes to complete. The professional organizations surveyed included the New York, Northeastern, North Central, South Central, and Southeastern sections of the American Urological Association; the European Society of

XSL•FO
RenderX

Residents in Urology; the European Association of Urology Section of Uro-Technology; Young Academic Urologists; the Spanish Urology Residents Working Group; the Sexual Medicine Society of North America; the Urological Society of Australia and New Zealand; and the Endourology Society. This project was approved by the IRB of the University of Miami (reference number: 20200414; approved 4/9/2020).

Statistical analysis was done using Microsoft Excel (Microsoft Corp) and MATLAB (version R2020a, The Mathworks Inc). We sought to compare our modified TUQ scores for the top three telehealth platforms used, as indicated by the survey results. The modified TUQ score was broken up into 5 categories for assessing telehealth platforms: usefulness, ease of use and learnability, interface quality, interaction quality, and reliability, with each question being answered based on a 7-point Likert scale. To calculate TUQ scores for each category, responses to all of the questions within each subsection of an individual's survey were averaged. Each category subsection was composed of 3 or 4 questions, based on the respective category. Next, all of the respondent's subsection scores, sorted by the telehealth platform used, were averaged to generate an overall subsection score stratified by platform used. These scores were compared using a single-factor ANOVA (analysis of variance), with an alpha of .05. Chi-squared tests were performed, with an alpha of .05. These results are expressed as total occurrences and corresponding percentages.

Results

Demographics

A total of 676 urologists clicked into the survey; however, after removing 56 incomplete surveys that were deemed to have insufficient data to be included, the final data set evaluated a total of 620 urologists from 58 different countries and 6 continents who participated in the survey. Of the 620 included, 24 were incomplete but were considered to have sufficient data to be included in the final data set. By continent, participation included 340 urologists from North America, 102 urologists from Europe, 65 urologists from Asia, 63 urologists from South America, 25 urologists from Africa, and 25 urologists from Australia. The three countries with the most participants were the United States (n=311, 50%), Spain (n=43, 6.9%), and Argentina (n=34, 5.5%). Further general demographics can be seen in Table 1. When evaluating participating urologist practices in terms of subspecialization, we divided practices into 4 groups: general (n=156, 25.2%), oncology (n=259, 41.8%), pediatrics (n=42, 6.8%), and nononcologists (n=163, 26.3%). Data on survey responses based on this subspecialty grouping are provided in Table 2. Overall, there were no significant differences in perceptions or usage of telemedicine among the subspecialty groups.

Table 1. Demographic data.

Variable	Respondents (N=620), n (%)
Gender	
Male	512 (82.6)
Female	101 (16.3)
Other	2 (0.3)
Prefer not to answer	5 (0.8)
Age (years)	
<30	23 (3.7)
30-39	183 (29.5)
40-49	183 (29.5)
50-59	129 (20.8)
≥60	102 (16.5)
Years of experience	
<5	144 (23.2)
5-10	119 (19.2)
11-15	98 (15.8)
16-20	73 (11.8)
>20	186 (30.0)
Subspecialty	
General	156 (25.2)
Oncology	259 (41.8)
Pediatrics	42 (6.8)
Nononcology	163 (26.3)
Continent	
Africa	25 (4.0)
Asia	65 (10.5)
Australia	25 (4.0)
Europe	102 (16.5)
North America	340 (54.8)

Table 2. Subspecialty grouped data.

Variable	Total, n (%)	General, n (%)	Oncology, n (%)	Pediatrics, n (%)	Nononcologists, n (%)	P value[a]
Clinic-anticipated inpatient consultation need						.49
<25%	144 (23.2)	33 (21.3)	59 (22.8)	9 (21.4)	43 (26.4)	
25%-49%	119 (19.2)	31 (19.9)	55 (21.2)	10(23.8)	23 (14.1)	
50%	98 (15.8)	23 (14.7)	41 (15.8)	10 (23.8)	24 (14.7)	
51%-75%	73 (11.8)	14 (9.0)	32 (12.4)	5 (11.9)	22 (13.5)	
>75%	186 (30)	55 (35.3)	72 (27.8)	8 (19.0)	51 (31.3)	
Surrogate use of telemedicine before COVID-19	302 (100)	22 (27.2)	43 (35.8)	9 (37.5)	24 (31.2)	.57
Have you ever used telemedicine?	620 (100)	81 (51.9)	121 (46.7)	24 (57.1)	76 (46.6)	.47
Clinic appointment using telemedicine prior to COVID-19						.57
<25%	83 (84.7)	17 (77.3)	36 (83.7)	8 (88.9)	22 (91.7)	
25%-49%	7 (7.1)	2 (9.1)	4 (9.3)	0 (0)	1 (4.2)	
50%	4 (4.1)	2 (9.1)	2 (4.7)	0 (0)	0 (0)	
51%-75%	1 (1.0)	1 (4.5)	0 (0)	0 (0)	0 (0)	
>75%	3 (3.1)	0 (0)	1 (2.3)	1 (11.1)	1 (4.2)	
Interest in using telemedicine for clinic appointments before COVID-19						.11
Not at all interested	65 (20.5)	23 (30.7)	20 (14.4)	2 (11.1)	20 (23.5)	
Not very interested	72 (22.7)	17 (22.7)	37 (26.6)	6 (33.3)	12 (14.1)	
Neutral	42 (13.2)	7 (9.3)	22 (15.8)	3 (16.7)	10 (11.8)	
Somewhat interested	86 (27.1)	19 (25.3)	34 (24.5)	6 (33.3)	27 (31.8)	
Very interested	52 (16.4)	9 (12)	26 (18.7)	1 (5.6)	16 (18.8)	
Surrogate use telemedicine since COVID-19	301 (100)	76 (93.8)	113 (95)	24 (100)	73 (94.8)	.68
Percentage of conversion to telemedicine since COVID-19						.39
<25%	86 (30.1)	25 (32.9)	36 (31.9)	10 (41.7)	15 (20.5)	
25%-49%	57 (19.9)	20 (26.3)	21 (18.6)	3 (12.5)	13 (17.8)	
50%	42 (14.7)	8 (10.5)	18 (15.9)	4 (16.7)	12 (16.4)	
51%-75%	34 (11.9)	9 (11.8)	12 (10.6)	4 (16.7)	9 (12.3)	
>75%	67 (23.4)	14 (18.4)	26 (23.0)	3 (12.5)	24 (32.9)	
Interest to continue telemedicine as surrogate after experience						.497
Not at all interested	11 (3.9)	4 (5.3)	2 (1.8)	1 (4.2)	4 (5.6)	
Not very interested	25 (8.8)	7 (9.2)	12 (10.8)	1 (4.2)	5 (6.9)	
Neutral	18 (6.4)	5 (6.6)	7 (6.3)	3 (12.5)	3 (4.2)	
Somewhat interested	82 (29)	29 (38.2)	27 (24.3)	7 (29.2)	19 (26.4)	
Very interested	147 (51.9)	31 (40.8)	63 (56.8)	12 (50.0)	41 (56.9)	
Telemedicine for interaction with hospital inpatients prior to COVID-19	37 (12.4)	13 (16.0)	15 (12.8)	2 (8.3)	7 (9.2)	.001
Telemedicine for interaction with hospital inpatient since COVID-19	55 (19.4)	16 (20.5)	24 (20.7)	7 (29.2)	8 (12.3)	.29
Percentage of conversion to telemedicine for inpatient visits						.27
<25%	19 (35.2)	9 (56.3)	8 (33.3)	0 (0)	2 (25)	

Variable	Total, n (%)	General, n (%)	Oncology, n (%)	Pediatrics, n (%)	Nononcologists, n (%)	P value[a]
25%-49%	11 (20.4)	2 (12.5)	5 (20.8)	3 (50)	1 (12.5)	
50%	7 (13.0)	2(12.5)	2 (8.3)	2 (33.3)	1(12.5)	
51%-75%	8 (14.8)	2 (12.5)	4 (16.7)	1 (16.7)	1 (12.5)	
>75%	9 (16.7)	1 (6.3)	5 (20.8)	0 (0)	3 (37.5)	
Interested in using telemedicine before?						.89
Not at all interested	25 (11.4)	7 (11.7)	8 (9.1)	2 (11.8)	8 (14.8)	
Not very interested	37 (16.9)	13 (21.7)	12(13.6)	4 (23.5)	8 (14.8)	
Neutral	41 (18.7)	12 (20.0)	18 (20.5)	2 (11.8)	9 (16.7)	
Somewhat interested	66 (30.1)	14 (23.3)	28 (31.8)	7 (41.2)	17 (31.5)	
Very interested	50 (22.8)	14 (23.3)	22 (25.0)	2 (11.8)	12 (22.2)	
Interested in using telemedicine after experience						.45
Not at all interested	3 (5.6)	0 (0)	3 (12.5)	0 (0)	0 (0)	
Not very interested	2 (3.7)	1 (6.3)	1 (4.2)	0 (0)	0 (0)	
Neutral	9 (16.7)	4 (25.0)	2 (8.3)	0 (0)	3 (37.5)	
Somewhat interested	20 (37.0)	7 (43.8)	9 (37.5)	2 (33.3)	2(25.0)	
Very interested	20 (37.0)	4 (25.0)	9 (37.5)	4 (66.7)	3 (37.5)	
Location of telemedicine						.18
At home	108 (38.2)	27 (35.1)	36 (32.4)	12 (52.2)	33 (45.8)	
In office	169 (59.7)	50 (64.9)	71 (64)	11 (47.8)	37 (51.4)	
At other places	6 (2.1)	0 (0)	4 (3.6)	0 (0)	2 (2.8)	
Interested in interacting with future patients with telemedicine after experience						.01
Not at all interested	7 (2.2)	3 (4)	2 (1.5)	0 (0)	2 (2.4)	
Not very interested	12 (3.8)	4 (5.3)	4 (3)	2 (11.1)	2 (2.4)	
Neutral	41 (13.1)	9 (12)	20 (14.9)	0 (0)	12 (14.1)	
Somewhat interested	122 (39.1)	36 (48)	58 (43.3)	9 (50)	19 (22.4)	
Very interested	130 (41.7)	23 (30.7)	50 (37.3)	7 (38.9)	50 (58.8)	

[a]Chi-squared test.

Barriers to Telemedicine Use

Approximately half (n=318, 51.2%) of the urologists surveyed have never used telemedicine. When assessing the barriers to telemedicine use, the top three reasons urologists gave were patients' lack of technological comprehension, patients' lack of access to required technology, and reimbursement concerns (Table 3). Another barrier of significance was lack of administrative support, which was the fourth most frequently mentioned barrier to telemedicine use. Of note, the question regarding barriers to usage had a "check all that apply" option, which explains why the percent of cases exceeds 100% in Table 3.

Table 3. Barriers to telemedicine use.

Reasons	Responses, n (%)	Cases (%)
Patients lack technological comprehension to use it	183 (18.8)	58.3
Patients lack access to necessary technology	174 (17.9)	55.4
Insurance reimbursement concerns	163 (16.7)	51.9
Lack of administrative support	142 (14.6)	45.2
Legal concerns	118 (12.1)	37.6
Practice lacks technology for telemedicine	118 (12.1)	37.6
Practice lacks finances for telemedicine	54 (5.5)	17.2
Other	22 (2.3)	7.0
Total	974 (100.0)	310.2

Telemedicine Use During the COVID-19 Pandemic

Prior to COVID-19, only 15.8% (n=98) of urologists surveyed were using telemedicine in their clinical practices. During the pandemic, however, 46.1% (n=283) of all urologists surveyed were using telemedicine in their clinical practices. Since the start of COVID-19, 50% (n=143) of participating urologists converted at least half of their originally scheduled in-person clinic visits to telemedicine. Despite this increase in usage, 68% (n=421) of all urologists surveyed believed that at least half of their clinic appointments required an in-person visit. Urologists with prior telemedicine experience were less likely to believe that 50% of their patients required in-person visits when compared to those without telemedicine experience (43.0% [n=130] vs 52.8% [n=168], P=.015). In the inpatient setting, 6% (n=37) of surveyed urologists had used telemedicine prior to COVID-19 to interact with urology hospital inpatients. During the pandemic, 8.9% (n=55) of participating urologists utilized telemedicine in the inpatient setting.

Among the participating urologists without telemedicine experience, interest in usage of telemedicine increased from 43.7% (n=139) to 80.8% (n=257) during COVID-19. After using telemedicine during the pandemic, 80.9% (n=244) of urologists surveyed were interested in continuing to use it in their practice. About half of participating urologists (n=116, 52.9%) were interested in utilizing telemedicine in the inpatient hospital setting. The majority of sampled urologists who had experienced inpatient telemedicine use during the pandemic were interested in continuing its usage (n=308, 74%), and about half (n=29, 53%) of the urologists who had not used it in that setting were interested in doing so.

Assessment of Telemedicine Platforms

The 5 most common platforms used by participating urologists for telemedicine visits were Zoom, Doxy.me, Epic, WhatsApp, and Skype. Telemedicine was mostly done either in the office (n=169, 59.7%) or at home (n=108, 38.2%). We compared the scores of the 3 most commonly used platforms (Zoom, Doxy.me, Epic) by usefulness, ease of use and learnability, interface quality, interaction quality, and reliability. There were no significant differences between the platforms in any category (Table 4).

Table 4. Usability scores of different telemedicine platforms.

Score and platform	Mean (SD)	F	F$_{critical}$	*df*	*P* value
Usefulness		1.82	3.10	2	.17
Overall	5.05 (1.24)				
Zoom	4.89 (1.31)				
Doxy.me	4.95 (1.28)				
Epic	5.67 (0.57)				
Ease of use and learnability		2.12	3.10	2	.12
Overall	5.46 (1.20)				
Zoom	5.51 (1.00)				
Doxy.me	5.04 (1.31)				
Epic	5.61 (0.76)				
Interface quality		0.03	3.10	2	.97
Overall	4.98 (1.17)				
Zoom	4.99 (1.11)				
Doxy.me	4.93 (1.23)				
Epic	4.93 (0.93)				
Interaction quality		0.86	3.10	2	.43
Overall	5.04 (1.14)				
Zoom	5.07 (1.11)				
Doxy.me	4.74 (1.26)				
Epic	5.07 (0.98)				
Reliability		0.76	3.10	2	.47
Overall	3.93 (1.16)				
Zoom	3.95 (1.25)				
Doxy.me	3.63 (1.17)				
Epic	3.73 (1.05)				

Assessment of individual TUQ questions provided insight into participating urologists' attitudes toward telemedicine use (Table 5). The majority of the urologists using telemedicine agreed that it improves patient access (n=223, 78.5%), saves physician travel time to the hospital (n=168, 55.6%), and addresses patients' health care needs (n=229, 80.6%). Evaluating ease of use, the majority of surveyed urologists found telemedicine simple to use (n=220, 77.5%), easy to learn (n=253, 89.1%), and felt they could be productive using it (n=213, 75%). Overall, 75.4% (n=214) of sampled urologists said they liked using telemedicine, but 57% (n=162) felt that telemedicine visits were not the same as in-person visits and 47.6% (n=135) did not think they could see their patients as well as if they were in person.

Table 5. Telehealth Usability Questionnaire survey data (N=284 responses).

Question	Strongly disagree, n (%)	Disagree, n (%)	Somewhat disagree, n (%)	Neutral, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)
Q25. TM[a] improves patient access to me	10 (3.5)	11 (3.9)	9 (3.2)	31 (10.9)	66 (23.2)	96 (33.8)	61 (21.5)
Q26. TM saves me traveling time to the hospital or clinic	18 (6.3)	47 (16.5)	11 (3.9)	50 (17.6)	44 (15.5)	71 (25)	43 (15.1)
Q27. TM provides for patient health care needs	6 (2.1)	10 (3.5)	20 (7.0)	19 (6.7)	82 (28.9)	110 (38.7)	37 (13)
Q28. TM was simple to use	8 (2.8)	14 (4.9)	25 (8.8)	17 (6.0)	49 (17.3)	113 (39.8)	58 (20.4)
Q29. TM was easy to learn to use	4 (1.4)	4 (1.4)	8 (2.8)	15 (5.3)	38 (13.4)	137 (48.2)	78 (27.5)
Q30. I could become productive quickly using TM	6 (2.1)	14 (4.9)	19 (6.7)	32 (11.3)	58 (20.4)	101 (35.6)	54 (19)
Q31. Interaction with TM is pleasant	7 (2.5)	7 (2.5)	17 (6)	26 (9.2)	71 (25)	115 (40.5)	41 (14.1)
Q32. I like using TM	6 (2.1)	13 (4.6)	26 (9.2)	25 (8.8)	57 (20.1)	109 (38.4)	48 (16.9)
Q33. TM is simple and easy to understand	4 (1.4)	7 (2.5)	4 (1.4)	25 (8.8)	54 (19)	141 (49.6)	49 (17.3)
Q34. TM is able to do everything I want it to do	31 (10.9)	52 (18.3)	55 (19.4)	26 (9.2)	62 (21.8)	42 (14.8)	16 (5.6)
Q35. I can easily talk with patients on TM	1 (0.4)	11 (3.9)	19 (6.7)	29 (10.2)	63 (22.2)	110 (38.7)	51 (18)
Q36. I can hear patient easily on TM	2 (0.7)	9 (3.2)	18 (6.3)	28 (9.9)	68 (23.9)	113 (39.8)	46 (16.2)
Q37. I was able to express myself effectively on TM	1 (0.4)	10 (3.5)	15 (5.3)	19 (6.7)	62 (21.8)	130 (45.8)	47 (16.5)
Q38. I can see the patient as well as if we met in person	24 (8.5)	49 (17.3)	62 (21.8)	25 (8.8)	59 (20.8)	44 (15.5)	21 (7.4)
Q39. Visits over TM are the same as in-person visits	44 (15.5)	73 (25.7)	45 (15.8)	25 (8.8)	59 (20.8)	24 (8.5)	14 (4.9)
Q40. I can recover quickly and easily from mistakes I made using TM	3 (1.1)	14 (4.9)	21 (7.4)	80 (8.2)	57 (20.1)	81 (28.5)	28 (9.9)
Q41. TM gave error signals that told me how to fix the problem.	23 (8.1)	68 (23.9)	25 (8.8)	105 (37)	34 (12)	24 (8.5)	5 (1.8)

[a]TM: telemedicine.

Discussion

Principal Findings

The COVID-19 pandemic has resulted in policies that have limited direct human interactions on a global scale. These limitations have drastically modified social and professional practices, especially within medicine. In an attempt to reduce transmission of COVID-19, there has been a large push for expansion of telemedicine usage throughout the medical field, including within urology. We wanted to explore how practicing urologists are adapting to using telemedicine during COVID-19, the usability of the telemedicine platforms, and the potential barriers to its use. Our results indicate that telemedicine was readily adopted during the pandemic as its usage among surveyed urologists almost tripled from 15.8% to 46.1%. Experiences during COVID-19 dramatically increased interest in telemedicine use among those participating urologists without

access to telemedicine (43.7% to 80.8%), and 80.9% of participating urologists using telemedicine were interested in continuing its usage. Despite the increased usage and interest in the continued usage of telemedicine during COVID-19, approximately half (51.2%) of the urologists surveyed had never used telemedicine, and the majority of urologists still believed that at least half of their clinic appointments require an in-person visit.

Prior to the pandemic, few urologists were using telemedicine due to lack of technological literacy, concerns over reimbursement, and resistance to changes in the workplace [3]. Our data confirms that the majority of urological practices were not utilizing telemedicine prior to the pandemic, as only 15.8% of sampled urologists were employing it in their practices. Despite the push for increased use of telemedicine during the pandemic, it is concerning that approximately half of participating urologists were still not using telemedicine.

XSL•FO
RenderX

Evaluating the top three barriers for telemedicine use, the main obstacle appears to be the technology itself. Technological issues, including disparities in internet access and poor audio quality in patients who speak non-English languages, have been described as barriers to telemedicine care during the pandemic [14]. A systematic review assessing the barriers to telemedicine adoption worldwide also confirmed that technology-specific issues were the main barriers to telemedicine use, with the most frequently cited barrier as technically challenged staff [15]. One possible way to help overcome some of these barriers could be through novel approaches using social media. Twitter-based journal clubs have been shown to be efficient means of disseminating information in a free and time-efficient manner [16]. Twitter-based educational strategies that educate health care providers, administrators, and patients on telehealth could facilitate previous apprehensions with the use of the technology and promote further usage. As telemedicine continues to play a major role in medical practices, hopefully health care systems and providers will invest more time and effort into educating patients on and providing better access to telemedicine.

Although only half of surveyed urologists were using telemedicine during the pandemic, those urologists with experience in telemedicine successfully incorporated it into their practices during quarantine. Health care systems were able to accommodate an increased telemedicine load since approximately 50% of participating urologists converted half of their originally scheduled in-person clinic visits to remote ones. A systematic review of telehealth in urology showed that telehealth can be successfully implemented in urology patients with prostate cancer, urinary incontinence, pelvic organ prolapse, uncomplicated urinary stones, and uncomplicated urinary tract infections, but the pandemic has forced urologists to assimilate virtually all urological conditions to telemedicine [17]. This rapid pivot to telemedicine in the outpatient setting is encouraging, demonstrating that urologists, patients, and health care systems alike are willing to provide and accept care over telemedicine. The prompt assimilation to telemedicine during the pandemic was well documented by the urology department at the tertiary academic center Charleston Area Medical Center. They demonstrated that quick telemedicine adoption was feasible; by day 5 of their transition, more patients were participating in audiovisual video encounters than any other modality. Just as important, no office staff found the introduction of telemedicine stressful, and more than 80% of telemedicine patients were interested in future encounters [18]. Despite these successes, our data still showed that lack of administrative support was a major barrier for many urologists to adopt telemedicine into their practice. With almost 15% of participating urologists citing it as a barrier, we need to acknowledge that there are overarching health system issues, not just patient and physician issues, that are preventing many from gaining access to telemedicine.

The success of transitioning to telemedicine is also dependent on patient acceptance of the technology. Data prior to COVID-19 already suggested patients are embracing telemedicine, but in the face of the pandemic, desire for telemedicine consultations as a replacement for in-person visits among urology patients was very high [19,20]. There are concerns about telemedicine and how it may introduce additional stressors to already vulnerable patients, especially the elderly who may lack technical knowledge or have disabilities such as hearing loss. In these cases, it is up to the urologist to accommodate these patients accordingly and ensure through their best practices that patients are engaging in clearly communicated encounters that includes shared decision making [21]. Although there is still significant room for improvement, the successful transition to telemedicine during the pandemic demonstrated health care systems can change rapidly and effectively when there is consensus between physicians, health care administrators, and patients.

Overall, telemedicine experiences during COVID-19 appear to have changed urologists' interest in future telemedicine usage. Urologists' experiences with telemedicine during the pandemic appears to have been positive with most participating urologists (80.9%) now interested in continuing to incorporate it in their practices. The majority of surveyed urologists using telemedicine agreed that it improves patient access (78.5%), saves physician travel time to the hospital (55.6%), and provides patients with their health care needs (80.6%). Most participants found telemedicine simple to use (77.5%), easy to learn (89.1%), and felt they could be productive using it (75%). Overall, 75.4% of urologists said they liked using telemedicine. The pandemic demonstrated the many benefits of telemedicine, which include reduced spread of COVID-19, reduced contamination of uninfected persons, reduced transmission to hospital workers and hospital surfaces, increased appointment convenience for patients, increased patient satisfaction, and provided physicians the opportunity to work from home [9]. Even though the majority of surveyed urologists wanted to incorporate telemedicine in their practice, 68% of participating urologists still believed that at least half of their clinic appointments required an in-person visit. Subanalysis of this population, however, demonstrated more acceptance of telemedicine as a replacement for in-person visits among those urologists using telemedicine. Urologists with prior telemedicine experience were less likely to believe that 50% of their patients required in-person visits when compared to those without telemedicine experience. Grouping participating urologists into general, oncology, pediatrics, and nononcology groups showed no significant differences in telemedicine practices or opinions. Our data suggest that even though participating urologists have experienced success with telemedicine, they still believed that most patients benefit from in-person visits. Further studies are needed to assess the advantages and disadvantages of telemedicine compared to in-person visits.

A major factor in the success of telemedicine is the platform itself. There is no universally accepted platform at this time, with different institutions adopting different programs. Our study was the first to investigate what telemedicine platforms were most commonly being used among urologists and to compare the usability of these platforms based on a modified version of the TUQ. The 3 most commonly used telemedicine platforms by our participants were Zoom, Doxy.me, and Epic. There were no significant differences between the platforms in any category. Of note, of the top 5 most used platforms, Zoom, WhatsApp, and Skype are not designed for telemedicine despite

being used for this purpose. Prior to COVID-19, all telemedicine platforms had to adhere to strict Health Insurance Portability and Accountability Act (HIPAA) specifications. With COVID-19 and the push for telemedicine, emergency provisions allowed for the usage of non-HIPAA compliant technologies like Facetime by Apple and WhatsApp. It is still recommended that when possible urologists utilize reliable and secure platforms that meet HIPAA standards [22]. Telemedicine is still fairly new, and although there are many upsides to its use, our data suggest that there is still much room for improvement for the technology to better accommodate current health care practices.

Limitations and Conclusion

There are several limitations to our study. As this study was distributed partly through social media, participation may be skewed to urologists who are more comfortable with computer technologies and therefore are more likely to use telemedicine in their practice. Although we were able to capture data on over 600 urologists, it is only a small fraction of the entire urology community and therefore may not be an accurate representation of the community. It also must be noted that the majority of participants were from North America and Europe, which share similar Western socioeconomic qualities compared to the less represented urologists from Asia, South America, and Africa. Therefore, our study may be considered a more accurate representation of telemedicine usage in Western countries. In addition, our use of telemedicine as defined by video calls may have excluded a large portion of health care provided by urologists that was performed over the phone. Video calls may not be well established in many countries and the role of telephone consultations was not assessed in our study. Strategies to improve the responses in online survey studies include sending reminders to respond, offering incentives to respond, and keeping the surveys short. Despite these limitations, this is the first study to quantify the use, usability, and pervading interest in telemedicine among urologists during the COVID-19 pandemic. In the face of this pandemic, urologists have demonstrated the ability to adopt and adapt telemedicine into their practices, but barriers involving the technology itself are preventing many from utilizing it despite increasing interest.

References

1. Voran D. Telemedicine and beyond. Mo Med 2015;112(2):129-135 [FREE Full text] [Medline: 25958658]
2. Kane CK, Gillis K. The Use Of Telemedicine By Physicians: Still The Exception Rather Than The Rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]
3. Castaneda P, Ellimoottil C. Current use of telehealth in urology: a review. World J Urol 2020 Oct 27;38(10):2377-2384. [doi: 10.1007/s00345-019-02882-9] [Medline: 31352565]
4. Teoh JY, Ong WLK, Gonzalez-Padilla D, Castellani D, Dubin JM, Esperto F, UroSoMe Working Group. A Global Survey on the Impact of COVID-19 on Urological Services. Eur Urol 2020 Aug;78(2):265-275 [FREE Full text] [doi: 10.1016/j.eururo.2020.05.025] [Medline: 32507625]
5. WHO Timeline - COVID-19. World Health Organization. 2020 Apr 27. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-05-15]
6. COVID-19: recommendations for management of elective surgical procedures. American College of Surgeons. 2020 Mar 13. URL: https://www.facs.org/covid-19/clinical-guidance/elective-surgery#:~:text=Reschedule%20elective%20surgeries%20as%20necessary,to%20care%20for%20these%20patients [accessed 2020-05-15]
7. Stensland KD, Morgan TM, Moinzadeh A, Lee CT, Briganti A, Catto JW, et al. Considerations in the Triage of Urologic Surgeries During the COVID-19 Pandemic. Eur Urol 2020 Jun;77(6):663-666 [FREE Full text] [doi: 10.1016/j.eururo.2020.03.027] [Medline: 32279903]
8. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
9. Rodriguez Socarrás M, Loeb S, Teoh JYC, Ribal MJ, Bloemberg J, Catto J, et al. Telemedicine and Smart Working: Recommendations of the European Association of Urology. Eur Urol 2020 Jul 09:2020 [FREE Full text] [doi: 10.1016/j.eururo.2020.06.031] [Medline: 32654801]
10. Grimes CL, Balk EM, Crisp CC, Antosh DD, Murphy M, Halder GE, et al. A guide for urogynecologic patient care utilizing telemedicine during the COVID-19 pandemic: review of existing evidence. Int Urogynecol J 2020 Jun 27;31(6):1063-1089 [FREE Full text] [doi: 10.1007/s00192-020-04314-4] [Medline: 32342112]

XSL•FO

RenderX

11.   Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016 Jul 01;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]

12.   Serwe KM. The Provider's Experience of Delivering an Education-based Wellness Program via Telehealth. Int J Telerehab 2018 Dec 11;10(2):73-80. [doi: 10.5195/ijt.2018.6268]

13.   Cheng O, Law N, Tulk J, Hunter M. Utilization of Telemedicine in Addressing Musculoskeletal Care Gap in Long-Term Care Patients. JAAOS Glob Res Rev 2020 Apr 14;4(4):e19.00128. [doi: 10.5435/jaaosglobal-d-19-00128]

14.   Watts KL, Abraham N. "Virtually Perfect" for Some but Perhaps Not for All: Launching Telemedicine in the Bronx during the COVID-19 Pandemic. J Urol 2020 Nov;204(5):903-904. [doi: 10.1097/JU.0000000000001185] [Medline: 32519903]

15.   Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]

16.   Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res 2015;17(4):e103 [FREE Full text] [doi: 10.2196/jmir.4194] [Medline: 25908092]

17.   Novara G, Checcucci E, Crestani A, Abrate A, Esperto F, Pavan N, Research Urology Network (RUN). Telehealth in Urology: A Systematic Review of the Literature. How Much Can Telemedicine Be Useful During and After the COVID-19 Pandemic? Eur Urol 2020 Jun 18 [FREE Full text] [doi: 10.1016/j.eururo.2020.06.025] [Medline: 32616405]

18.   Davis C, Novak M, Patel A, Davis C, Fitzwater R, Hale N. The COVID-19 Catalyst: Analysis of a Tertiary Academic Institution's Rapid Assimilation of Telemedicine. Urology Practice 2020 Jul;7(4):247-251. [doi: 10.1097/upj.0000000000000155]

19.   Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]

20.   Boehm K, Ziewers S, Brandt MP, Sparwasser P, Haack M, Willems F, et al. Telemedicine Online Visits in Urology During the COVID-19 Pandemic-Potential, Risk Factors, and Patients' Perspective. Eur Urol 2020 Apr 27 [FREE Full text] [doi: 10.1016/j.eururo.2020.04.055] [Medline: 32362498]

21.   Sosnowski R, Kamecki H, Joniau S, Walz J, Klaassen Z, Palou J. Introduction of Telemedicine During the COVID-19 Pandemic: A Challenge for Now, an Opportunity for the Future. European Urology 2020 Jul. [doi: 10.1016/j.eururo.2020.07.007]

22.   Gadzinski AJ, Gore JL, Ellimoottil C, Odisho AY, Watts KL. Implementing Telemedicine in Response to the COVID-19 Pandemic. Journal of Urology 2020 Jul;204(1):14-16. [doi: 10.1097/ju.0000000000001033]

Abbreviations

ANOVA: analysis of variance
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
TUQ: Telehealth Usability Questionnaire

XSL•FO
RenderX

Review

Barriers and Enablers in Implementing Electronic Consultations in Primary Care: Scoping Review

Rebecca Baines[1], MSc; John Tredinnick-Rowe[1], PhD; Ray Jones[1], PhD; Arunangsu Chatterjee[1], PhD

University of Plymouth, Plymouth, United Kingdom

Corresponding Author:
Arunangsu Chatterjee, PhD
University of Plymouth
Room 108 I 3 Endsleigh Place
Drake Circus
Plymouth,
United Kingdom
Phone: 44 1752 586837
Email: arunangsu.chatterjee@plymouth.ac.uk

Abstract

Background: Often promoted as a way to address increasing demands, improve patient accessibility, and improve overall efficiency, electronic consultations are becoming increasingly common in primary care, particularly in light of the current COVID-19 pandemic. However, despite their increasing use, a theoretically informed understanding of the factors that support and inhibit their effective implementation is severely limited.

Objective: With this scoping review, we sought to identify the factors that support and inhibit the implementation of electronic consultations in primary care.

Methods: In total, 5 electronic databases (PubMed, Medline, Embase, CINAHL, and PsycINFO) were systematically searched for studies published in 2009-2019 that explored the impact and/or implementation of electronic consultations in primary care. Database searches were supplemented by reference list and grey literature searches. Data were analyzed using inductive thematic analysis and synthesized using Normalization Process Theory (NPT).

Results: In total, 227 articles were initially identified and 13 were included in this review. The main factors found to hinder implementation included awareness and expectations; low levels of engagement; perceived suitability for all patient groups, conditions, and demographics; cost; and other contextual factors. Reports of information technology reliability and clinical workload duplication (as opposed to reduction) also appeared detrimental. Conversely, the development of protocols and guidance; patient and staff education; strategic marketing; and patient and public involvement were all identified as beneficial in facilitating electronic consultation implementation.

Conclusions: This review highlights the need for proactive engagement with patients and staff to facilitate understanding and awareness, process optimization, and delivery of coherent training and education that maximizes impact and success. Although the necessity to use online methods during the COVID-19 pandemic may have accelerated awareness, concerns over workload duplication and inequality of access may remain. Future research should explore health inequalities in electronic consultations and their economic impacts from multiple perspectives (eg, patient, professional, and commissioner) to determine their potential value. Further work to identify the role of meaningful patient involvement in digital innovation, implementation, and evaluation is also required following the rapid digitization of health and social care.

Introduction

With a 16% increase in the number of general practitioner (GP) consultations between 2007-2014 in England alone, primary care is considered to be at "saturation point" [1]. General practice is often described as facing increasing demand, reduced accessibility, and heightened patient dissatisfaction [2,3]. As a result, primary care providers are being increasingly encouraged to adopt alternative, more digitally focused methods of care

provision [4], particularly in light of the recent COVID-19 outbreak. This drive toward digital platforms is reflected on an international scale [5-8], with digital platforms often promoted as a way to relieve pressures on existing services and improve accessibility, efficiency, and cost-effectiveness while simultaneously promoting self-management and patient-centered care [1,3,4,9]. However, critical exploration of these assumptions is severely limited.

Within the United Kingdom, prior to the COVID-19 pandemic, there were two main electronic consultation providers: askmyGP and eConsult, the latter of which was previously known as WebGP [10,11]. These systems are online triage tools designed to provide patients with an alternative way of contacting their GP practice [4]. For the purposes of this article, we define the term electronic consultations as an online service that enables patients to access advice and care from a primary care practitioner or staff member by combining preliminary health issues and symptom checking with appointment booking. These products are distinct from video-based consultations that a doctor might use, such as Babylon, AccuRx, or LIVI. In the United Kingdom, electronic consultations are also distinct from their use in North America as a video system to communicate between family doctors and specialists in hospitals.

While some research has explored the benefits of electronic consultations and the experiences of health care professionals using them, such work is often critiqued for its overreliance on pilot studies [3,4]. Other criticisms of existing literature also include a limited understanding of the cost implications for both health care professionals and patients [12-14], and a predominate focus on professional perspectives [1,15], with limited exploration of patient experiences and expectations [3,6,11]. Furthermore, of the limited research conducted, the majority of it has focused on the experience of using electronic consultations, as opposed to the practicalities of implementing such technology. Finally, in spite of their increasing use, a theoretically informed understanding of the factors that support or inhibit the implementation of electronic consultations is severely limited [4,5,8,11], particularly in a UK setting [16]. This review seeks to address this gap by identifying the factors that support and inhibit the implementation of electronic consultations in primary care using Normalization Process Theory (NPT) as a theoretical framework. This scoping review is well-timed given the rapidly increasing use of electronic consultations in light of the COVID-19 outbreak and the resulting mandatory shift toward total triage in primary care [4]. Although focused on general practice and electronic consultations, the implications of this review may be relevant to other digital forms of health and care technology and their subsequent implementation.

Methods

Design

A scoping review was conducted due to their ability to map existing evidence in an emerging field (such as electronic consultations), identify gaps in existing understanding, and incorporate different study designs that can be grouped together to evaluate a particular topic of interest [17,18].

Search Strategy

As advised by an information specialist, 5 databases (PubMed, Medline, Embase, CINAHL, and PsycINFO) were systematically searched using the search terms "econsult" OR "electronic consultation" OR "WebGP" OR "non face-to-face consultations" OR "technology mediated consultations" AND "primary care" OR "GP" OR "general practice." Search terms were designed and reviewed using the Peer Review of Electronic Search Strategies (PRESS) guidance [19].

A grey literature search was also conducted to ensure sufficient inclusivity and coverage. Grey literature was defined as "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers" [20]. The peer-reviewed search strategy was also used in Google Scholar.

Inclusion and Exclusion Criteria

Research studies that explored the impact or implementation of electronic consultation platforms designed to be used by a primary care clinician and patient, published in the English language, of any study design (including opinion pieces and editorial letters) were included. Research studies that were not published in the English language or focused on electronic consultation platforms outside of primary care (eg, between non–primary care specialties) were excluded to retain a relevant focus of interest.

While the researchers aimed to be inclusive, due to limited resources, a sensitive translation of non-English texts could not be provided. Finally, to ensure only the most contemporary literature was included, a time limit was applied (January 1, 2009, to January 31, 2019). Literature searches were conducted on February 1, 2019.

Screening and Eligibility

All identified articles were screened using a two-stage process. First, the title and abstract of all identified articles were reviewed independently by two researchers using the inclusion criteria outlined above. This process was facilitated using Rayyan, software for conducting reviews [21]. The full texts of potentially relevant articles were then reviewed for inclusion. Any disagreements were resolved through the inclusion of a third researcher. Database searches were also supplemented by reference list searches of included studies.

Data was extracted from the included studies independently by two reviewers using a pilot tested extraction form. Data extracted included author names, date of publication, setting, study type, sample, analysis method, and reported findings/author interpretations.

Analysis

Included studies were initially coded independently by two reviewers using inductive thematic analysis as outlined by Braun and Clarke [22]. Identified themes were then synthesized using NPT as outlined below by the same two reviewers.

Synthesis

Data was synthesized using NPT [23] as an analytical framework. NPT was originally developed to understand the

embedding of new, particularly complex technologies in health care systems [23], providing a clear rationale for its inclusion in this research. As outlined by Murray [23], NPT is underpinned by four constructs that often operate simultaneously:

1. Coherence: how people make sense of a new technology/system/process.
2. Cognitive participation: how people engage with a new technology/system/process.
3. Collective action: how people make new technologies/systems/processes work in practice (or not).
4. Reflexive monitoring: how people assess the value of new technologies/systems/processes.

For the purposes of this article, we report findings on the first three domains of NPT (coherence, participation, and collective action). As previously mentioned, of the limited research previously conducted, most has focused on professional experiences (eg, reflexive monitoring). We therefore focus on the three remaining domains of NPT to avoid duplication.

Quality Appraisal

In line with scoping review practice [18], included studies were not quality appraised.

Ethical Approval

Participation in this research was entirely voluntary. All participants gave full informed consent. The University of Plymouth, Faculty of Health and Human Sciences (Reference number 18/19-1060) provided ethical approval.

Results

Overview

From the 227 articles initially identified, 13 were included for the purposes of this review (Figure 1). Table 1 summarizes the characteristics of the included studies.

Figure 1. PRISMA diagram of included studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses.
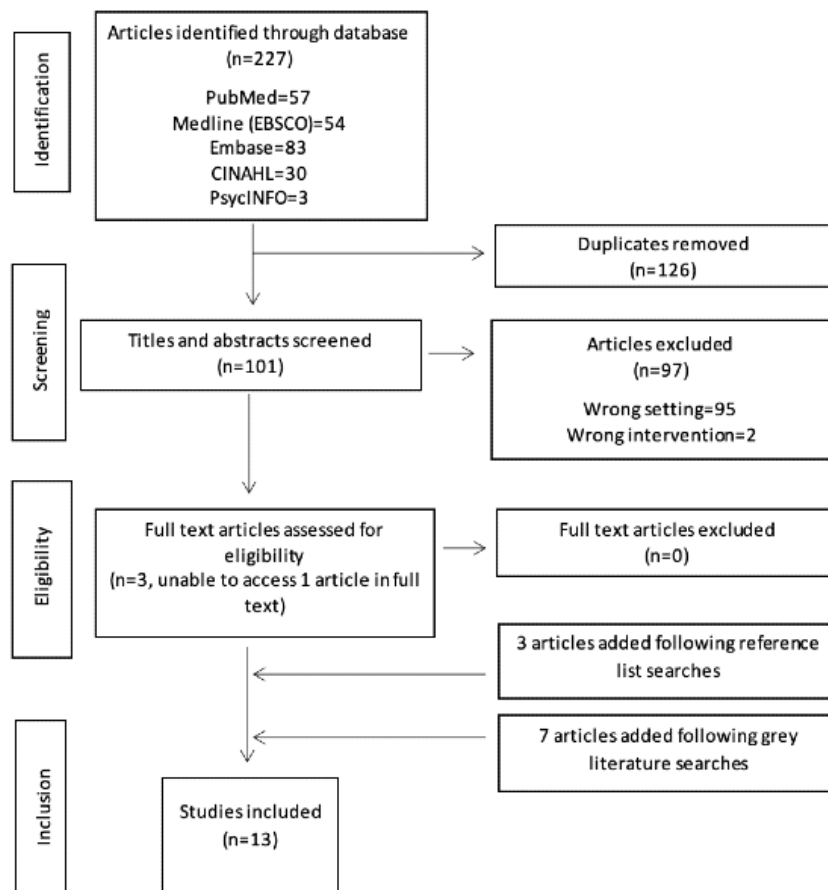
Table 1. Characteristics of included studies.

Authors	Country	Study type	Intervention	Setting	Participants and databases	Analysis method
Atherton et al, 2018 [24]	England and Scotland	Mixed methods case design	Alternatives to face-to-face consultations	General practices with varied experience of implementing alternatives to face-to-face consultations	Patients and practice staff	Descriptive statistics, multivariate analysis, and coding of qualitative data from EMIS Health
Banks et al, 2018 [4]	West of England	Qualitative	eConsult	General practices that piloted an e-consultation system for 15 months during 2015 and 2016	23 semistructured interviews with staff members	Thematic analysis of interview data
Cowie et al, 2018 [25]	Scotland	Mixed methods	eConsult	11 general practices piloting eConsult	44 semistructured practice staff interviews, 1 focus group (4 staff), and 291 patient survey responses	Health economics, thematic analysis and coding, patient survey, descriptive statistics
Carter et al, 2018 [6]	South West England	Mixed methods	WebGP	6 GP practices in Devon	Six practices provided consultations data; 20 GPs completed case reports (regarding 61 e-consults); 81 patients completed questionnaires; 5 GPs and 5 administrators were interviewed	Statistical analysis of surveys, thematic analysis of interview data
Farr et al, 2018 [3]	South West England	Mixed methods	eConsult	6 general practices	23 practice staff interviews, patient survey data for 756 e-consultations from 36 practices	Economic analysis on usage and costs, Normalization Process Theory, inductive thematic analysis, patient survey and record data statistical analysis
Ogden, 2018 [26]	N/A[a]	Opinion piece	Online consulting	N/A	N/A	Descriptive statistics
Marshall et al, 2018 [8]	N/A	Opinion piece	Online consulting	N/A	N/A	N/A
Casey et al, 2017 [1]	N/A	Mixed methods case study	Online consultation system	Inner-city general practice	Information technology developers, clinicians, and administrative staff	Thematic and discourse analysis of interview data
Edwards et al, 2017 [16]	South West of England	Observational study	eConsult	South West of England	eConsult data obtained from 36 general practices	Economic analysis, website analytics, survey-based statistical analysis
Wise, 2017 [10]	N/A	Opinion piece	Online consultations	N/A	N/A	N/A
Hanna et al, 2012 [27]	Scotland	Qualitative	Non–face-to-face consultations	Scotland	20 semistructured interviews with general practitioners	Thematic analysis of interviews
Mair et al, 2012 [28]	N/A	Systematic review	eHealth systems (only information pertaining to online consultations recorded)	N/A	MEDLINE, EMBASE, CINAHL, PsycINFO, and the Cochrane Library were searched for reviews published between January 1, 1995, and March 17, 2009. Studies had to be systematic reviews, narrative reviews, qualitative meta-syntheses or meta-ethnographies of eHealth implementation	Evidence synthesis

Authors	Country	Study type	Intervention	Setting	Participants and databases	Analysis method
Hanna et al, 2011 [15]	Scotland	Mixed methods	Non–face-to-face consultations	Scotland	600 practice manager questionnaire responses, 20 practice manager interviews	Chi-squared test from a survey, thematic analysis of interviews

[a]N/A: not applicable.

As demonstrated in Table 1, most of the research published in the last two years has used a mixed methods design. There has been a particular focus on the introduction of electronic consultations in South West England and Scotland. Our results are presented in line with the three selected domains of NPT as outlined above. We begin by presenting the factors that inhibit implementation, followed by those that facilitate implementation.

Factors That Inhibit Electronic Consultation Implementation

Coherence: What Factors Inhibit the Understanding of Electronic Consultations?

Awareness and Expectations

A lack of understanding and awareness about the purpose of electronic consultation systems appears particularly problematic in their implementation and impact [24]. Patient understanding and awareness was reportedly mixed, with the effectiveness of electronic consultation advertising often called into question [16,25]. Patients were often not aware of alternative methods available to them, and how to access or operate such services [16,24]. Atherton suggests this is a matter for individual practices rather than policy directives [24].

A difference in staff and patient expectations regarding the appropriate use of electronic consultations was also reported [3]. This primarily related to the inappropriate use of electronic consultations to get a timelier face-to-face appointment as later discussed.

Cognitive Participation: What Factors Inhibit How People Engage With Electronic Consultations?

Low Uptake

Studies identified a low uptake as a further barrier to electronic consultation implementation [1,3,4,6,11,16,24]. Engagement levels were often reported to be much lower than expected [3]. For example, Banks et al [4] reported that only around 6% of practices used some form of electronic consultation. Similarly, Edwards et al [16] concluded that of the 36 practices reviewed, there was an average of 2 electronic consultations per 1000 patients per month. Such findings mirrored calculations by Cowie and colleagues [11]; the highest submission rate reported was almost 0.1 consultations per patient per year.

Suitability for Certain Patient Groups, Conditions, and Demographics

The suitability of electronic consultations for certain patient groups, conditions, and demographics was also called into question. Electronic consultations were considered more suitable for certain patient groups with discrete, "straightforward," or familiar ongoing treatment queries [1,3,24,27]. In comparison, face-to-face consultations were reportedly preferred for new, acute, or complex cases that required physical examination and/or regular monitoring [3,4,16,24,26]. As a result, some health care professionals expressed concern that patient groups including those with long-term conditions, multimorbidity, and/or mental health problems would be disadvantaged or unintentionally excluded from electronic consultation opportunities [8,24]. Some GPs repeatedly expressed concerns that introducing a new technology could create or widen existing inequalities in access to health care [3,27,29]. Atherton acknowledged the purposeful selection of patients by health care professionals, including receptionists and administrators, based on their assumptions of who they felt would be able to use electronic consultations appropriately [24], highlighting further concerns regarding unintentional or intentional exclusion.

In regard to demographics, included articles also suggested that women are almost twice as likely to use electronic consultation systems than men [6,11,16]. Younger patients are also more likely to engage with electronic consultation systems, with levels of engagement typically declining with age [3,6,11,16,24]. Levels of education, language [3], and employment were also shown to influence electronic consultation use [6,24,27].

Contextual Factors: Practice Size, Deprivation, and Geographical Location

Following this, a range of contextual factors including practice size, deprivation, and geographical location appeared to influence electronic consultation implementation and effect [29]. Hanna et al [29] suggested that the bigger the practice size, the more support available, including a website, an information technology (IT) manager, and a triage team, thereby influencing implementation efforts and sustainability [29].

Cost

Finally, cost was reported as a significant barrier to electronic consultation engagement and implementation [3,4], with limited evidence available to justify its associated expense [4,11]. As suggested by Farr et al [3], costs often prohibited long-term engagement with electronic consultation systems.

Collective Action: What Factors Inhibit How People Work With Electronic Consultations?

Additional Time and Increased Workload

In regard to how people work with electronic consultations (collective action), the majority of included studies reported an increase in clinician workload. This was often attributed to additional follow-up and/or face-to-face consultations due to a lack of sufficient information being provided [1,4,6,11,16,27,28]. Although one study acknowledged a potential benefit of having the clinical issue documented prior to a face-to-face appointment [4], this conclusion was

XSL·FO

RenderX

contradicted by another article that found that follow-up appointments were typically longer than the national average (14.5 minutes versus 9.2 minutes for face-to-face) [16]. However, it is important to note that this disparity may be a result of a lack of GP continuity between electronic and follow-up consultations [3].

GPs also reported significant difficulties in establishing a patient's primary concern, following an inability to probe for further information [3,4,16]. Where reported, most electronic consultations resulted in either follow-up phone calls or face-to-face appointments, leading GPs to report a duplication in their workload, as opposed to a desired reduction [1,3,6]. The relatively low uptake of electronic consultation systems as previously described appeared to exacerbate this concern [11].

An increase in administrative workload was also reported, particularly the frequent need to contact patients regarding their electronic consultation outcome or to arrange a face-to-face or telephone appointment [11]. Cowie et al [11] reported that most practices made at least three attempts to contact patients, frequently reporting that contact could not be made. The use of a withheld number was suggested as a possible explanation. Some practices have developed methods to facilitate email contact as a solution to this problem, although issues of data security and protection remained a concern [11].

Medicolegal Concerns

Linked to concerns of data security and protection, medicolegal issues were also identified as a barrier to electronic consultation implementation and use. Concerns about the negotiation of clinical risk and diagnosis uncertainty were repeatedly expressed within the context of an increasingly litigious culture [8,27,29].

Potential to "Game" the System

As previously mentioned, some practice staff members expressed concern regarding "patients gaming the system" [4], with some reportedly using the system to achieve a timelier face-to-face appointment [3,4,6,11]. As reported by Banks et al [4] and others, staff often felt patients could get an appointment quicker using electronic consultation methods by circumventing traditional, often telephone-based appointment systems. The responsibility of the practice to contact the patient within a specified time frame, often by the end of the next working day [4,6], as well as the content of marketing material [11], appeared to influence this perception.

IT Reliability

Finally, the reliability of technology was seen as a potential barrier to electronic consultation implementation and use. As reported by Hanna et al [29], some interviewees felt new IT systems were highly reliable, while others expressed concern about network availability and speed [27,29]. The interoperability of IT systems was also identified as influential in electronic consultation implementation [3,11], with competing or conflicting IT systems proving to be problematic.

Factors That Facilitate Implementation

In addition to the barriers outlined above, a number of facilitating factors were also identified in regard to electronic consultation implementation and use. Each are discussed in turn below.

Coherence: What Factors Facilitate the Understanding of Electronic Consultations?

Protocols and Strategies

First, the provision of protocols, strategies, and/or guidance including medicolegal advice was considered integral to the effective implementation of electronic consultations [11,26-28]. As identified by Cowie et al [11], the development of a clear strategy for introducing electronic consultations prior to any implementation was considered fundamental. Such protocols should be developed in collaboration with staff, patients, and electronic consultation champions (as discussed below) [11].

Education

Following this, quality patient and staff education on what electronic consultations are, what they are not, and when to use them was widely encouraged [3,6,11]. Cowie et al [11] recommended the construction of practice process diagrams along with clear objectives to facilitate staff understanding and awareness. An exploration of staff expectations was also considered integral [11].

Focused Marketing

Linked to the provision of education was a desire for more focused marketing [6,11,26]. Identified ways of promoting and marketing electronic consultations included leaflets, clear website positioning, posters, and recorded telephone messages [26]. As suggested by Ogden et al [26], recorded messages appeared particularly influential when recorded by a doctor.

Patient and Public Involvement and Wider Network Development

Combining the development of protocols, education, and marketing, patient and public involvement was also considered integral prior to and during any electronic consultation implementation to ensure acceptability and understanding [3,11]. Despite this, few practices reported engaging with patients in electronic consultation design, implementation, or evaluation [3]. This may explain some of the issues identified regarding patient understanding and awareness. Patient participation groups (PPGs) were identified as a beneficial resource for facilitating patient engagement, as were patient and public involvement workshops to provide feedback on electronic consultation systems, protocols, and experiences [11].

Cognitive Participation: What Factors Support How Stakeholders Engage With Electronic Consultation?

Staff Training

Moving on to cognitive participation, how stakeholders engaged with electronic consultations, adequate staff training that addressed how electronic consultations operate, how electronic consultations fit with existing practice processes, and individual responsibilities were considered essential for successful implementation [8,11,26-28].

eConsultation Champions

An electronic consultation champion was also considered helpful in ensuring effective implementation and engagement by promoting its use among patients and more reticent staff members [11,28]. However, the risk of jeopardizing staff and patient commitment by recruiting a critical champion was also acknowledged [28].

Collective Action: What Factors Facilitate How Stakeholders Work With Electronic Consultations?

Strategic Marketing

With regard to collective action (how stakeholders work with electronic consultations), strategic marketing or signposting was identified as a facilitating factor [3,6,11]. For example, signposting patients to use electronic consultations in certain situations where only remote GP access was likely to be required, including follow-up appointments, general administrative queries, repeat prescriptions, and general advice [11]. A more focused marketing strategy was considered to help improve overall efficiency, ensuring patients who might benefit from the service most were directly encouraged to engage with it [11].

Notification Alerts

In response to concerns of engaging in telephone "ping-pong," a desire for the development of an alerting system to inform patients of an incoming call following an electronic consultation request was also expressed [6,11]. Some practices already provide an estimated time for patient call-backs, clearly informing the patient that this may be from an unavailable or unknown number [11]. The development and effective incorporation of email contact was also expressed as a desirable solution to improving electronic consultation implementation and effectiveness [11], although this is reliant on the smooth integration of electronic consultations into existing IT systems as described below [28].

Integration of Technology and Adequate Resourcing

Mair et al [28] suggested clinicians may be deterred or become resistant to working with a system that adds complexity or requires additional effort and time. IT support was therefore considered integral to implementation [27], as was adequate resourcing, particularly financial support [28].

Pairing of GPs With Patients

Finally, the possibility of allocating electronic consultations to GPs who had had previous contact with the patient was seen as an effective way to facilitate implementation [3]. Table 2 provides a visual representation of the barriers and facilitators outlined in this review, according to the relevant domains of NPT.

Table 2. Identified barriers and facilitators to electronic consultation implementation and use.

Normalization Process Theory domain	Barriers	Facilitators
Coherence	• Lack of understanding regarding its purpose and intended use	• Development of protocols, strategies, and guidance, including medicolegal advice • Patient and staff education • Focused marketing • Wider consultation with patients and staff members prior to implementation
Participation	• Low uptake • Mainly administrative requests • Suitability for certain patient groups and conditions • Contextual factors including practice size, deprivation, and geographical location • Cost • Limited patient involvement	• Staff training • Strategic/targeted patient use for those most likely to benefit from electronic consultations • Electronic consultation champion
Action	• Purposeful patient selection. • Additional time or increased workload • Medicolegal concerns • The potential to "game" the system • Information technology reliability	• Effective signposting informing patients of when to use electronic consultation and when not • Notification alerts to alleviate administrative issues related to contacting patients • Integration of technology and adequate resourcing • Pairing of general practitioner with prior patient contact

Discussion

Summary

This review addressed an identified gap in existing literature by developing a theoretically informed understanding of the factors that support and inhibit the practical implementation of electronic consultations in primary care [4,5,8,11,16]. Review findings suggest limited staff and patient awareness and understandings; low levels of engagement; perceived suitability for all patient groups, conditions, and demographics; cost; and other contextual factors including practice size, levels of deprivation, and geographical location are the main inhibitory factors of effectively implementing electronic consultations. As a result, the majority of included studies reported an increase in clinician and admin time, with many GPs reporting a duplication in workload, as opposed to the desired reduction. Findings from our review also identified a number of factors

that could help facilitate the effective implementation of electronic consultations. These primarily focused on the provision of staff training, protocols, strategies, and guidance; enhanced patient awareness and education; strategic marketing; notification alerts, and patient and public involvement in the innovation, implementation, and evaluation stages.

Comparison With Existing Literature

Similar to existing research, the included articles reported a shortcoming of electronic consultations in their current form [4], with expressed skepticism regarding their financial investment [3]. Furthermore, many health care professionals expressed concern that electronic consultation duplicated administrative and clinical workloads. Such findings have been widely reported by a number of studies [1,3,4,6,16]. However, it is important to note that this may be due to the absence of supporting factors identified in this review and the relatively new emergence of electronic consultations. Further exploration of whether the presence of facilitating factors improves electronic consultation would be beneficial. Finally, we found a limited amount of research exploring electronic consultation experiences and impact from a patient perspective [3,6,11], identifying a further area for future research.

Strengths and Limitations

Strengths of this review include its exploration of both peer-reviewed and grey literature and novel application of a theoretical framework in the context of general practice and electronic consultations. However, its limitations must also be acknowledged. In line with scoping review practice, included articles were not quality appraised. The exclusion of non–English language texts may also have introduced research bias. Future research may benefit from a wider range of bibliographic databases, including technical databases such as IEEE. Exploration of any differences between private and free at the point of access service implementations/commissioning would also be beneficial.

Implications for Practice

With these in mind, the implications from this review are clear. First, the implementation of electronic consultations appears most effective when both patients and staff members are involved in the design, implementation, and evaluation of their processes and outcomes. Second, the rationale, purpose, and intended use of electronic consultations needs to be effectively communicated to both patients and staff members to ensure appropriate use and implementation. This could be best achieved through targeted marketing as well as meaningful patient involvement to facilitate patient understanding and acceptability. Third, marketing materials should reflect the reality of the product proposed to effectively manage people's expectations (ie, electronic consultation may not reduce workload to the extent originally promised). Efforts should also be made wherever possible to incorporate the factors identified as beneficial in electronic consultation implementation to achieve maximum success and impact. Finally, further work is required to explore the potential inequalities in electronic consultations, with evidence to suggest some patient groups may be disadvantaged or unintentionally excluded from electronic consultation opportunities [8,23]. Further exploration of the economic costs and benefits of electronic consultations from the perspective of patients, professionals, and commissioners would also be beneficial in informing current debates.

Conclusion

In conclusion, the implementation of electronic consultations in primary care can be facilitated by the development of protocols and strategies, patient and staff education, accurate and targeted education, and meaningful patient and public involvement. Efforts should be made wherever possible to incorporate factors identified as beneficial in facilitating electronic consultations to ensure maximum impact and success. Further research exploring the economic impacts of electronic consultations would be beneficial from a patient, professional, and commissioner perspective.

Conflicts of Interest

None declared.

References

1. Casey M, Shaw S, Swinglehurst D. Experiences with online consultation systems in primary care: case study of one early adopter site. Br J Gen Pract 2017 Oct 09;67(664):e736-e743. [doi: 10.3399/bjgp17x693137]
2. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
3. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Dec 19;8(3):e019966. [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]

4. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8. [doi: 10.3399/bjgp17X693509] [Medline: 29109115]

5. Atherton H, Ziebland S. What do we need to consider when planning, implementing and researching the use of alternatives to face-to-face consultations in primary healthcare? Digit Health 2016 Nov 21;2:2055207616675559. [doi: 10.1177/2055207616675559] [Medline: 29942570]

6. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):e018688. [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]

7. Joschko J, Keely E, Grant R, Moroz I, Graveline M, Drimer N, et al. Electronic Consultation Services Worldwide: Environmental Scan. J Med Internet Res 2018 Dec 21;20(12):e11112 [FREE Full text] [doi: 10.2196/11112] [Medline: 30578187]

8. Marshall M, Shah R, Stokes-Lampard H. Online consulting in general practice: making the move from disruptive innovation to mainstream service. BMJ 2018 Mar 26;360:k1195. [doi: 10.1136/bmj.k1195] [Medline: 29581174]

9. Johnston DL, Murto K, Kurzawa J, Liddy C, Keely E, Lai L. Use of Electronic Consultation System to Improve Access to Care in Pediatric Hematology/Oncology. J Pediatr Hematol Oncol 2017 Oct;39(7):e367-e369. [doi: 10.1097/MPH.0000000000000833] [Medline: 28437292]

10. Wise J. Electronic consultations offer few benefits for GP practices, says study. BMJ 2017 Nov 06:j5141. [doi: 10.1136/bmj.j5141]

11. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a Digital Consultation and Self-Care Advice Tool in Primary Care: A Multi-Methods Study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]

12. Liddy C, Keely E, Afkham A. Faster Access, Better Care: A Guide to Establishing an eConsult Service. 2016. URL: https://ruor.uottawa.ca/bitstream/10393/34968/1/Faster%20Access%2c%20Better%20Care.pdf [accessed 2020-10-28]

13. Anderson D, Villagra VG, Coman E, Ahmed T, Porto A, Jepeal N, et al. Reduced Cost Of Specialty Care Using Electronic Consultations For Medicaid Patients. Health Aff (Millwood) 2018 Dec;37(12):2031-2036. [doi: 10.1377/hlthaff.2018.05124] [Medline: 30633678]

14. Anderson D, Villagra V, Coman EN, Zlateva I, Hutchinson A, Villagra J, et al. A cost-effectiveness analysis of cardiology eConsults for Medicaid patients. Am J Manag Care 2018 Jan 01;24(1):e9-e16 [FREE Full text] [Medline: 29350511]

15. Hanna L, May C, Fairhurst K. The place of information and communication technology-mediated consultations in primary care: GPs' perspectives. Fam Pract 2012 Jun;29(3):361-366. [doi: 10.1093/fampra/cmr087] [Medline: 22006040]

16. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901. [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]

17. Osman MA, Schick-Makaroff K, Thompson S, Featherstone R, Bialy L, Kurzawa J, et al. Barriers and facilitators for implementation of electronic consultations (eConsult) to enhance specialist access to care: a scoping review protocol. BMJ Open 2018 Sep 28;8(9):e022733. [doi: 10.1136/bmjopen-2018-022733] [Medline: 30269069]

18. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]

19. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 2009 Sep;62(9):944-952. [doi: 10.1016/j.jclinepi.2008.10.012] [Medline: 19230612]

20. GreyNet. 4th International Conference on Grey Literature, New Frontiers in Grey Literature, 4-5 October 1999, Washington DC. Asian Libraries 1999 Jul;8(7):1. [doi: 10.1108/al.1999.17308gab.010]

21. Rayyan - a web and mobile app for systematic reviews. URL: https://rayyan.qcri.org/ [accessed 2020-10-28]

22. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

23. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]

24. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. Health Serv Deliv Res 2018 Jun;6(20):1-200. [doi: 10.3310/hsdr06200] [Medline: 29889485]

25. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a Digital Consultation and Self-Care Advice Tool in Primary Care: A Multi-Methods Study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]

26. Ogden J. Online consultations: the future of NHS GP appointments? Prescriber 2018 Sep 19;29(9):27-30. [doi: 10.1002/psb.1703]

27.   Hanna L, May C, Fairhurst K. The place of information and communication technology-mediated consultations in primary care: GPs' perspectives. Fam Pract 2012 Jun 17;29(3):361-366. [doi: 10.1093/fampra/cmr087] [Medline: 22006040]
28.   Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364. [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
29.   Hanna L, May C, Fairhurst K. Non-face-to-face consultations and communications in primary care: the role and perspective of general practice managers in Scotland. Inform Prim Care 2011;19(1):17-24 [FREE Full text] [doi: 10.14236/jhi.v19i1.789] [Medline: 22118332]

Abbreviations

GP: general practitioner
IT: information technology
NPT: Normalization Process Theory

XSL·FO
RenderX

Review

Effects of Different Telemonitoring Strategies on Chronic Heart Failure Care: Systematic Review and Subgroup Meta-Analysis

Hang Ding[1,2,3*], PhD; Sheau Huey Chen[4*], PG; Iain Edwards[5], MBA; Rajiv Jayasena[6], PhD; James Doecke[6], PhD; Jamie Layland[7,8], MBChB, MD, PhD; Ian A Yang[9], MBBS, PhD, FRACP, FAPSR, FThorSoc, Grad Dip Clin Epid; Andrew Maiorana[4,10], PhD

[1]RECOVER Injury Research Centre, Faculty of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia

[2]The Australian e-Health Research Centre, Commonwealth Scientific & Industrial Research Organisation, Brisbane, Australia

[3]Prince Charles Hospital - Northside Clinic Unit School, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[4]School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia

[5]Department of Community Health, Peninsula Health, Melbourne, Australia

[6]The Australian e-Health Research Centre, Commonwealth Scientific & Industrial Research Organisation, Melbourne, Australia

[7]Department of Cardiology, Peninsula Health, Melbourne, Australia

[8]Peninsula Clinical School, Monash University, Melbourne, Australia

[9]Department of Thoracic Medicine, The Prince Charles Hospital, The University of Queensland, Brisbane, Australia

[10]Allied Health Department and Advanced Heart Failure and Cardiac Transplant Service, Fiona Stanley Hospital, Perth, Australia

[*]these authors contributed equally

Corresponding Author:
Andrew Maiorana, PhD
School of Physiotherapy and Exercise Science
Curtin University
GPO Box U1987
Perth, 6845
Australia
Phone: 61 8 9266 9225
Email: A.Maiorana@curtin.edu.au

Abstract

Background: Telemonitoring studies in chronic heart failure are characterized by mixed mortality and hospitalization outcomes, which have deterred the uptake of telemonitoring in clinical practice. These mixed outcomes may reflect the diverse range of patient management strategies incorporated in telemonitoring. To address this, we compared the effects of different telemonitoring strategies on clinical outcomes.

Objective: The aim of this systematic review and subgroup meta-analysis was to identify noninvasive telemonitoring strategies attributing to improvements in all-cause mortality or hospitalization outcomes for patients with chronic heart failure.

Methods: We reviewed and analyzed telemonitoring strategies from randomized controlled trials (RCTs) comparing telemonitoring intervention with usual care. For each strategy, we examined whether RCTs that applied the strategy in the telemonitoring intervention (subgroup 1) resulted in a significantly lower risk ratio (RR) of all-cause mortality or incidence rate ratio (IRR) of all-cause hospitalization compared with RCTs that did not apply this strategy (subgroup 2).

Results: We included 26 RCTs (N=11,450) incorporating 18 different telemonitoring strategies. RCTs that provided medication support were found to be associated with a significantly lower IRR value than RCTs that did not provide this type of support (P=.01; subgroup 1 IRR=0.83, 95% CI 0.72-0.95 vs subgroup 2 IRR=1.02, 95% CI 0.93-1.12). RCTs that applied mobile health were associated with a significantly lower IRR (P=.03; IRR=0.79, 95% CI 0.64-0.96 vs IRR=1.00, 95% CI 0.94-1.06) and RR (P=.01; RR=0.67, 95% CI 0.53-0.85 vs RR=0.95, 95% CI 0.84-1.07).

Conclusions: Telemonitoring strategies involving medication support and mobile health were associated with improvements in all-cause mortality or hospitalization outcomes. These strategies should be prioritized in telemonitoring interventions for the management of patients with chronic heart failure.

XSL•FO
RenderX

Introduction

Chronic heart failure (CHF) is a severe chronic disease [1] affecting over 26 million people worldwide [2]. Despite advances in modern medical therapy [3] and multidisciplinary clinical care [4], CHF continues to manifest a poor quality of life [5], frequent hospitalizations [6,7], low survival rates [8], and high health care expenditure [2]. Telemonitoring has been extensively studied as an innovative approach to enable care providers to remotely monitor patients at home and provide timely intervention in the event of clinical deterioration. Over the past two decades, many enabled care programs have been developed and evaluated, and several reviews have demonstrated the potential of using telemonitoring interventions to reduce mortality [9-11] and hospitalizations [9,11] in CHF care. However, the outcomes from individual randomized controlled trials (RCTs) are heterogeneous, with nonsignificant effects obtained in several large and well-designed RCTs [12-14]. Owing to these mixed outcomes, the use of telemonitoring in CHF care has been questioned [15,16] and has not yet been embraced in clinical recommendations [17,18].

Mixed outcomes in telemonitoring studies have been attributed to insufficient support from cardiologists, unsatisfactory patient compliance [19,20], low predictive power for clinical deterioration [14], and improvements in usual care [14,18]. However, these findings were limited to narrative analyses of individual telemonitoring studies. Several reviews have evaluated specific approaches to CHF care, including mobile health (mHealth) [21-23]; structured telephone [11,18], videophone, and interactive voice response devices [24]; education alone; pharmacist interventions; and clinical support by various care providers [25]. These reviews provide valuable insight into the effectiveness of specific types of interventions, but do not explain the mixed outcomes across telemonitoring interventions involving different components of care.

To address the existing knowledge gap, we conducted a systematic review and meta-analysis using a novel approach of evaluating the effect of different noninvasive telemonitoring strategies on reduced all-cause mortality and hospitalization to identify which strategies were associated with these outcomes.

Methods

Literature Search

This review was performed according to the Cochrane Collaboration methodological guidelines [26]. We conducted a literature search in the PubMed, EMBASE, CINAHL, and Cochrane Library databases, covering the publication period from January 1990 to February 2020. The publications were required to be (1) relevant to telehealth, telemedicine, telemonitoring, telecare, internet, mobile, smartphone, remote monitoring, or home monitoring; (2) involving patients with CHF; and (3) in the English language. An information specialist officer at the Commonwealth Scientific and Industrial Research Organization (Brisbane, Australia) and an expert librarian at Curtin University, Western Australia, Australia helped develop the bibliographies and conduct the database search (for a more complete description of our search strategy, see Multimedia Appendix 1).

Two investigators (HD and SC) independently reviewed the articles obtained. Disagreements between the two investigators were resolved by a third reviewer (AM or IE).

Scope of Telemonitoring

In this review, we employed a hierarchical structure considering that telehealth encompasses telemonitoring, as well as eHealth care processes and communication, telemedicine, and mHealth [27]. We then defined the scope of telemonitoring as "the use of information technology to monitor patients at a distance," as described by Meystre [28]. Finally, we included a telemonitoring intervention in the analysis if it involved "the transfer of physiological data such as blood pressure, weight, electrocardiographic signals, or oxygen saturation through technology such as telephone lines, broadband, satellite, or wireless networks" [27].

Inclusion and Exclusion Criteria

This review focused on noninvasive telemonitoring interventions evaluated through an RCT. The inclusion criteria were: (1) studies evaluating telemonitoring for CHF for at least 3 months, (2) prospective RCTs comparing telemonitoring-based care with usual care, and (3) full peer-reviewed journal articles reporting outcomes of all-cause mortality or all-cause hospitalization. The exclusion criteria were: (1) articles reporting preliminary analysis outcomes; (2) studies with a sample size less than 50 (Multimedia Appendix 2), because, compared with large studies, small studies are often associated with a lower level of reporting quality [29], are more likely to be heterogeneous [30], and overestimate outcome effects [31]; and (3) telemonitoring via implantable devices, as these interventions often involve a different care paradigm to noninvasive devices and have been the subject of dedicated reviews [32,33].

Telemonitoring Strategies Extracted

We extracted 18 telemonitoring strategies according to three categories: technology applications (6 strategies), care objectives (7 strategies), and care support methods (5 strategies) (Table 1).

Table 1. Extracted telemonitoring strategies for the subgroup meta-analysis on telemonitoring interventions for chronic heart failure (CHF).

Strategies	Descriptions
Technology applications	
mHealth[a] system (or combining with mHealth apps)	An mHealth system was used in the telemonitoring program, and the system involved a set of software apps mainly designed for mobile devices such as smartphones, personal digital assistants, and tablet computers.
PC[b]-based system	A PC-based system was used in the telemonitoring program, which involved a set of software apps mainly designed for PCs.
Weight scale	A device enabling participants to measure body weight and transfer the data to care providers in the telemonitoring program.
Blood pressure monitor	A device enabling participants to measure blood pressure and transfer the data to care providers in the telemonitoring program.
ECG[c] monitoring device	A device enabling participants to record ECG and transfer the data to care providers in the telemonitoring program.
Heart rate monitor	A device enabling participants to measure heart rate and transfer the data to care providers in the telemonitoring program.
Care objectives	
Education	The telemonitoring program included a care objective/component involving CHF education. The education content could be provided via video clips, animation, or text messages.
Daily weight monitoring	The telemonitoring program contained a care objective/component to assist the participants in daily weight monitoring. The assistance was delivered predominantly via automated messages and telephone calls.
Diet	The telemonitoring program contained a care objective/component for improving dietary behavior recommended for CHF.
Medication support	Clinical support was provided to optimally adjust medication therapy or support participants to adhere to the medication recommendations for CHF.
Exercise	Exercise was monitored or assessed via electronic questionnaires in the program. Clinical interventions such as automated messages and telephone calls were provided to assist participants in conducting exercises according to clinical recommendations.
Depression and anxiety	A care objective/component was specifically provided to address depression and anxiety in participants through the telemonitoring program.
Monitoring symptoms	Participants used telemonitoring apps to record their CHF-related symptoms. Accordingly, care providers reviewed the recorded symptoms and provided interventions.
Care support methods	
Collaborative care	Interventions and support for collaborative care were provided in the telemonitoring program, such as collaborative reviews, referrals, and communication for follow up.
Physician support	Physicians were included in the telemonitoring program to provide clinical intervention to the participants.
Nurse support	Nurses were included in the telemonitoring program to provide clinical intervention to the participants.
Call center support	A call center was included in the telemonitoring program to provide support to the participants.
Automated system	Automated systems were used to automatically monitor the participants' data and provide reminders, alerts, and notifications to the participants.

[a]mHealth: mobile health.

[b]PC: personal computer.

[c]ECG: electrocardiogram.

Review Outcomes

The risk ratio (RR) of all-cause mortality and the incidence rate ratio (IRR) of all-cause hospitalization in the RCTs were analyzed. The RR and IRR values in each RCT were calculated from the event counts of mortality and hospitalization. For each strategy, we divided the RCTs into two subgroups: RCTs that applied the strategy in the telemonitoring intervention (subgroup 1) and RCTs that did not apply the strategy (subgroup 2). We then compared the two subgroups (subgroup 1 vs subgroup 2), and examined whether the difference between the two groups in the RR and IRR outcomes was statistically significant.

Meta-Analysis

In the meta-analysis, we used a random-effects model with the DerSimonian–Laird estimator [34,35], and report the RR, IRR, and 95% CI for each group. For RCTs with no events in one arm, we applied a continuity correction of 0.5. The heterogeneity

of RCTs in each subgroup was examined by the Q test and I^2 statistic [36,37]. The statistical significance of heterogeneity was determined by a relaxed P value of .10 ($P_H < .10$) [38]. The I^2 values of 25%, 50%, and 75% were used to reflect a low, moderate, and high level of heterogeneity, respectively [37]. To evaluate the risk of bias, a regression test was used to analyze the asymmetry of a funnel plot of the RR or IRR results in a subgroup [36]. The regression test was used to examine whether the outcomes of individual RCTs were related to the corresponding sampling variances [39]. A significant regression ($P_F < .05$) indicated a significant risk of bias. The difference between the two groups (subgroup 1 vs subgroup 2) was evaluated by a Wald-type test [36], and statistical significance was determined if the corresponding two-sided P value was less than .05 ($P_C < .05$). A mixed-effects model [36] was also used to evaluate the effects of potential confounders, including sex, age, or the severity measure of left ventricular ejection fraction (LVEF). The meta-analysis methods and tests were performed using RStudio Version 1.1.383 [40] associated with the "metafor" meta-analysis package (version 2.0) [36].

Risk of Bias

A summary of the methodological risk of bias of the included studies was conducted in accordance with the Cochrane Handbook for Systematic Reviews of Interventions [26] by two investigators (HD, SC) using the risk of bias tool in the Cochrane Collaboration's review-writing software RevMan 5.3. This involved reporting the following individual elements for the included RCTs: random sequence generation, allocation sequence concealment, blinding of participants and personnel,

blinding of outcome assessment, completeness of outcome data, and selective outcome reporting. Each item was judged as being at a high, low, or unclear risk of bias. Studies were deemed to be at the highest risk of bias if they were scored at a high or unclear risk of bias for either the sequence generation or allocation concealment domains [26].

Results

Search Results

The literature search results are presented in Figure 1. We found 3870 records from the bibliographic search and 56 records from three existing systematic reviews [9,11,41] and a manual search, resulting in a total of 3926 records. After removing duplicates, we obtained 1632 articles for screening. In the screening process, we excluded 1553 articles because of absence of inclusion criteria and consequently obtained 79 articles for a full-text assessment. We then excluded 53 articles according to the inclusion and exclusion criteria, and one article because of its poor completion rate recognized by the authors [42]. Finally, this review included 26 RCTs. Among them, 25 RCTs provided all-cause hospitalization events and 21 RCTs provided mortality events.

Among the assessment elements of bias risk, the blinding of participants and personnel was the least used method in the RCTs included (Figure 2). There were 11 RCTs that did not blind participants and personnel (Figure 3). Nine RCTs did not report their blinding status and only six RCTs used a blinding approach. The blinding of outcome assessment was the least reported element, and 14 RCTs (54%) had "unclear risk of bias."

Figure 1. Flow diagram of study selection.
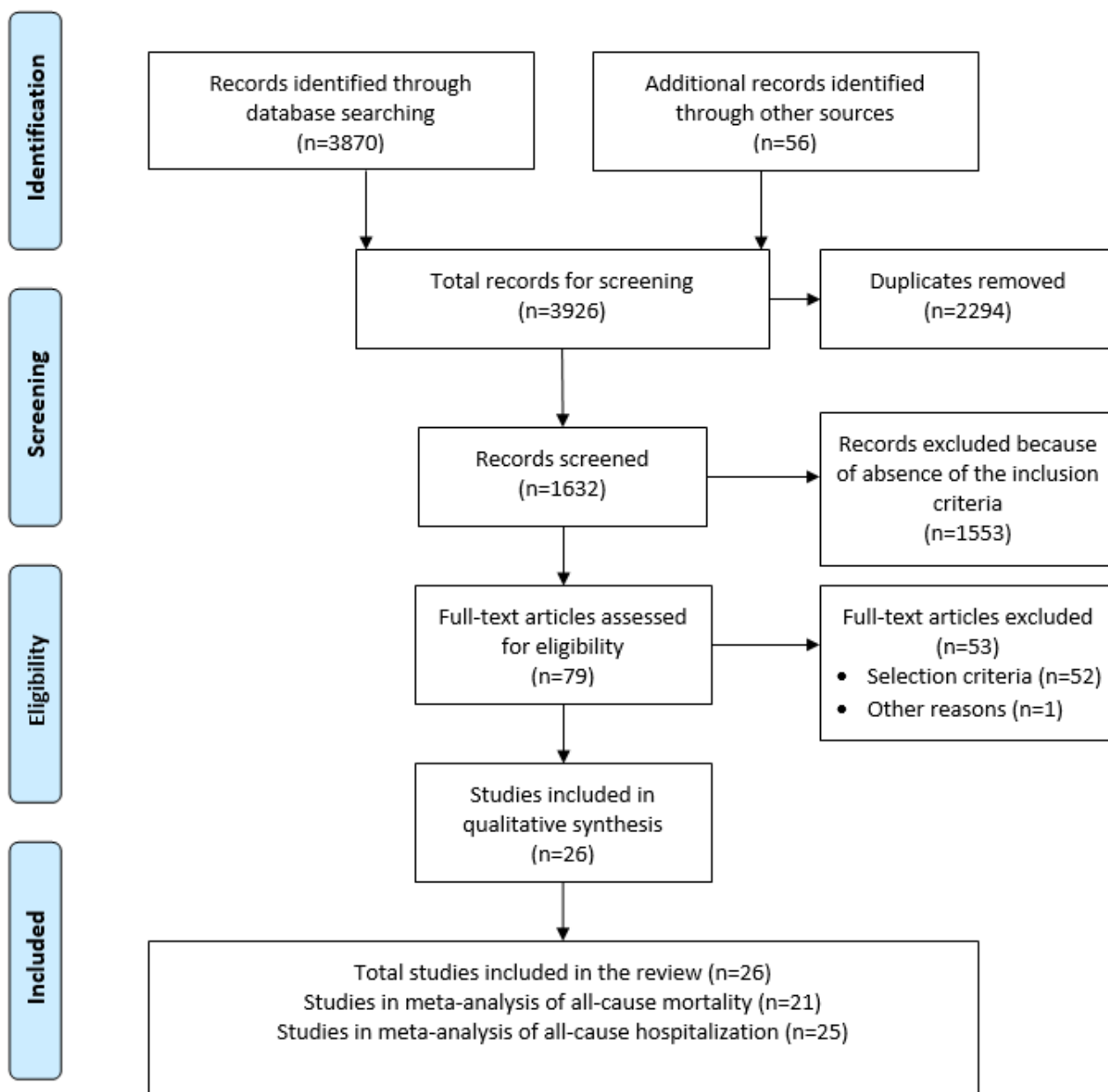
Figure 2. Risk of bias assessment. Authors' judgments about each methodological quality item are presented as percentages across all included studies.
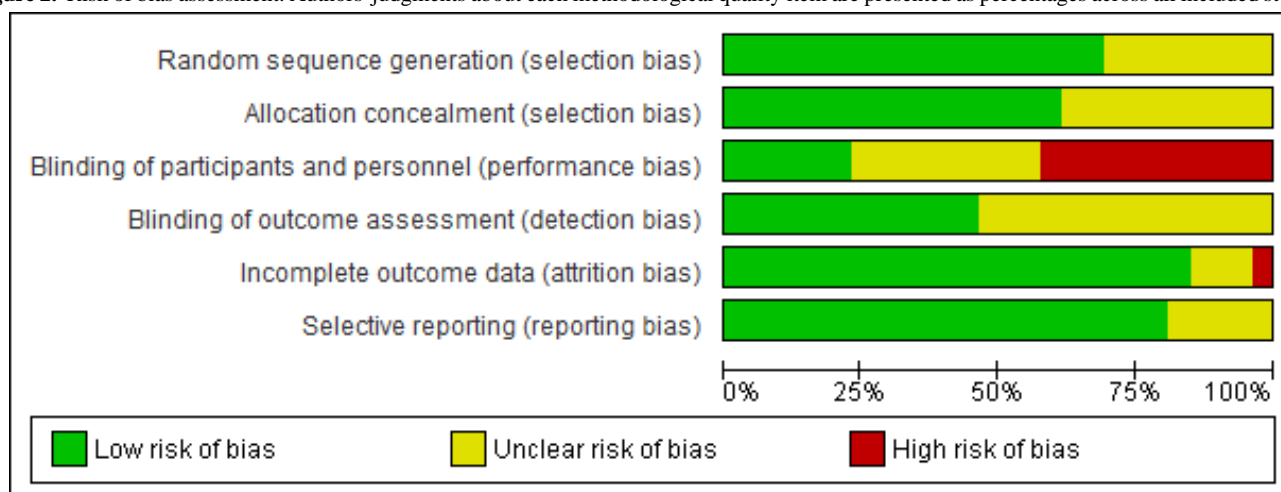
Figure 3. Risk of bias summary. Authors' judgements about each risk of bias item are summarized for each included study.

Participant Characteristics

The 26 RCTs included 11,450 participants. The participants' characteristics are shown in Table 2. The median age was 67.4 years and the median rate of male participants was 73.15% (8376/11,450). The participants generally had a significantly reduced (<40%) LVEF, with a median LVEF of 29.6%, and they experienced mild to moderate levels of symptoms, with a median New York Heart Association functional class score of 2.6. The median trial size of the RCTs was 290 participants and the median follow-up duration was 12 months.

Table 2. Participants' characteristics in 26 randomized controlled trials included in the subgroup meta-analysis.

Characteristic	Median (IQR)
Age (years)	67.40 (65.08-72.75)
Trial size (N)	290 (180-675)
Follow-up duration (months)	12 (6-12)
Male (%)	73.15 (66.00-79.95)
LVEF[a] (%)	29.60 (27.00-35.93)
NYHA[b] class score	2.6 (2.3-2.8)

[a]LVEF: left ventricular ejection fraction.

[b]NYHA: New York Heart Association.

Telemonitoring Strategies

We extracted 18 telemonitoring strategies from the 26 RCTs, as shown in Table 3. Some strategies were commonly used, such as telemonitoring weight scales (26/26, 100% RCTs), call-center support (24/26, 92%), and daily weight monitoring (25/26, 96%). Strategies that were not commonly used included nurse support (2/26, 8%), intervention for depression and anxiety (3/26, 12%), and exercise (3/26, 12%). The telemonitoring programs in the RCTs generally contained multiple strategies, with a mean of 8.7 strategies per care program.

Table 3. Telemonitoring strategies and randomized controlled trials included in the meta-analysis.

Reference	N	Care support method					Care objective							Technology application					
		Alerts	Nurse	Call Center	Physician	Collaborative	Ed[a]	Weight	Diet	Meds[b]	Ex[c]	DA[d]	Symptoms	PC[e] app	mHealth[f]	Scale	BP[g]	HR[h]	ECG[i]
[43]	384	0	0	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0
[44]	216	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	1	0
[45]	156	0	0	1	0	0	0	1	0	1	0	0	0	0	1	1	1	1	0
[46]	133	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0
[12]	1653	0	0	1	1	0	1	1	0	0	0	0	1	0	0	1	0	0	0
[47]	248	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	1
[48]	178	1	0	1	1	1	0	1	0	1	0	0	1	0	1	1	1	1	0
[49]	182	0	0	1	1	0	0	1	0	1	0	0	1	0	0	1	1	0	0
[50]	160	0	1	1	1	1	0	1	1	0	1	0	0	0	1	1	1	1	0
[51]	460	0	1	1	1	0	1	1	1	1	0	0	1	0	1	1	1	0	1
[52]	280	0	0	1	1	0	0	1	0	0	0	0	1	0	1	1	0	0	0
[53]	72	1	0	1	1	0	1	1	0	1	0	0	1	0	1	1	0	0	0
[54]	1360	0	0	1	1	1	0	1	1	1	0	0	0	1	0	1	1	1	0
[55]	1538	0	0	1	1	1	1	1	0	1	0	0	1	0	1	1	1	1	1
[13]	710	0	0	1	1	0	0	1	0	0	0	0	1	0	1	1	1	0	1
[56]	181	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0
[57]	319	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
[58]	72	1	0	1	1	0	1	1	1	1	0	0	1	0	1	1	0	0	0
[59]	261	0	0	1	1	0	1	1	0	0	0	0	1	0	1	1	1	1	1
[60]	339	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	1
[14]	1437	0	0	1	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0
[61]	100	1	0	1	1	0	0	1	0	1	0	0	1	0	1	1	1	0	1
[62]	315	0	0	1	1	1	1	1	0	1	0	0	1	0	0	1	0	0	0
[63]	80	0	0	0	1	0	0	1	0	0	0	1	1	0	1	1	1	0	1
[64]	316	0	0	1	0	0	1	1	0	1	1	0	0	0	0	1	1	0	0
[65]	300	1	1	0	0	1	1	1	0	1	0	0	1	0	0	1	1	1	0
Total	11,450	7	3	24	21	7	12	25	6	16	3	2	17	2	13	26	20	14	8

[a]Ed: education.

[b]Meds: medication.

[c]Ex: exercise.

[d]D/A: depression and anxiety.

[e]PC: personal computer.

[f]mHealth: mobile health.

[g]BP: blood pressure.

[h]HR: heart rate.

[i]ECG: electrocardiogram.

Overall Effectiveness of Telemonitoring

There were 21 RCTs (n=10,536) with event counts of all-cause mortality and 25 RCTs (n=9912) with event counts of all-cause hospitalization. The outcomes of mortality (RR) and hospitalization (IRR) with 95% CIs are shown in Figure 4 and Figure 5, respectively. Overall, telemonitoring interventions were found to be more effective than usual care on reducing

XSL•FO
RenderX

all-cause mortality (RR=0.85, 95% CI 0.76-0.94, P=.01) and all-cause hospitalizations (IRR=0.90, 95% CI 0.81-0.99, P=.04). The outcomes of both RR and IRR were heterogeneous (P_H=.001), with a low-to-moderate level of heterogeneity (I^2=35.3%) in the RR outcomes and a moderate-to-high level

of heterogeneity (I^2=73.2%) in the IRR outcomes. In the funnel plot–based test, the risk of bias was significant for both RR (z=1.89, P_F=.001 and IRR (z=3.33, P_F=.001) outcomes. We also used the mixed-effects model to adjust for sex, age, or LVEF, but did not find significant differences in these results.

Figure 4. Event counts and effectiveness of telemonitoring interventions on all-cause mortality. There were 20 randomized controlled trials (N=10,263) with mortality event counts in the subgroup meta-analysis. RR: relative risk.
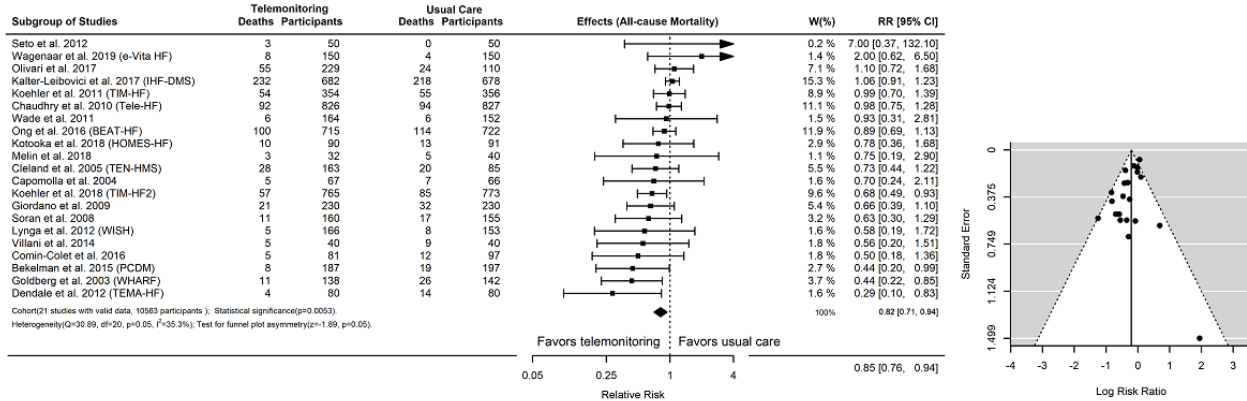
Figure 5. Event counts and effectiveness of telemonitoring interventions on all-cause hospitalization. There were 24 randomized controlled trials (N=9612) with hospitalization event counts in the subgroup meta-analysis. IRR: incidence rate ratio.
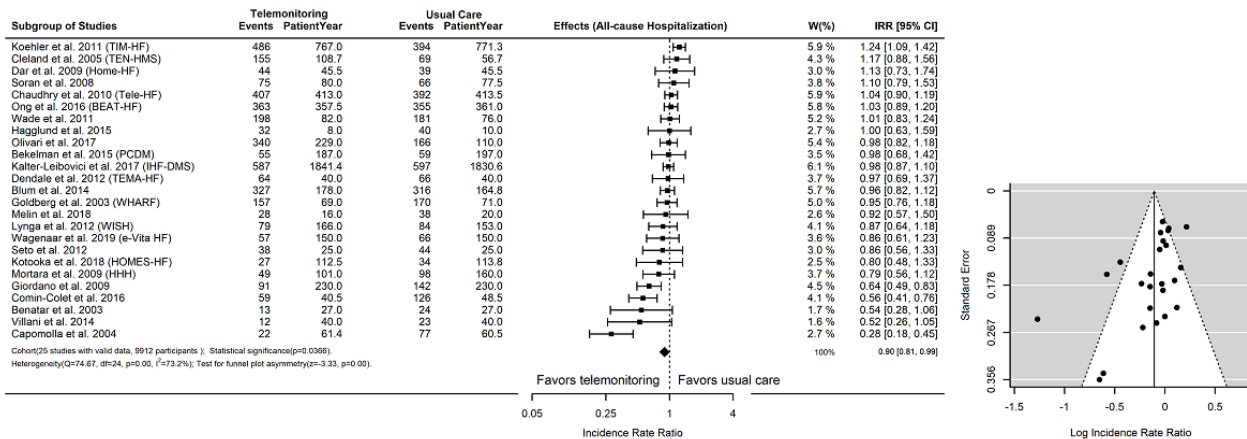
Subgroup Comparison of Telemonitoring Strategies

The subgroup of RCTs that provided medication support (subgroup 1, 15 RCTs, n=4563, IRR=0.83, 95% CI 0.72-0.95) was found to be associated with a significantly (P=.01) lower IRR of all-cause hospitalization than the comparison subgroup of RCTs that did not apply this strategy (subgroup 2, 10 RCTs, n=5349, IRR=1.02, 95% CI 0.93-1.12), as shown in Tables 4-6 and Multimedia Appendix 3. Within the subgroup of RCTs that provided medication support, the telemonitoring interventions were found to be more effective than usual care on reducing hospitalizations (15 RCTs, n=4563, IRR=0.83, 95% CI 0.72-0.95, P=.01). The IRR outcomes in both subgroups were heterogeneous (Multimedia Appendix 3 and Multimedia

Appendix 4). The outcomes in the comparison subgroup of RCTs that did not apply this strategy were associated with the risk of bias.

Similarly, the subgroup of mHealth (subgroup 1, 12 RCTs, n=2662, IRR=0.79, 95% CI 0.64-0.96) was associated with a significantly (P=.03) lower IRR of all-cause hospitalization than the comparison subgroup (subgroup 2, 13 RCTs, n=7250, IRR=1.00, 95% CI 0.94-1.06), as shown in Tables 4-6 and Multimedia Appendix 5. Within the mHealth subgroup, the telemonitoring interventions were found to be significantly more effective than usual care on reducing all-cause hospitalizations (subgroup 1, 12 RCTs, n=2662, IRR=0.79, 95% CI 0.64-0.96, P=.03). The IRR outcomes in the mHealth subgroup were heterogeneous and were associated with a risk of bias.

Table 4. Subgroup meta-analysis to examine the effect of telemonitoring strategies on all-cause hospitalization for randomized controlled trials (RCTs) that applied the strategy in the telemonitoring intervention (subgroup 1).

Strategies	RCTs (N participants)	Effect		Heterogeneity		Funnel test Z (P value)
		IRR[a] (95% CI)	P value	Q (P value)	I^2	
Technology application						
PC[b]-based system	2 (1493)	0.54 (0.16-1.81)	.32	25.07 (<.001)	96.0%	0.00 (<.001)
Blood Pressure Monitor	19 (7201)	0.87 (0.77-0.98)	.02	72.50 (<.001)	75.2%	−3.30 (.001)
ECG[c] Monitor	7 (2198)	0.91 (0.73-1.12)	.37	27.56 (<.001)	78.2%	−2.01 (.04)
Telemonitoring Weight Scale	25 (9912)	0.90 (0.83-0.99)	.03	74.67 (<.001)	67.9%	−3.24 (.001)
Heart Rate Monitor	13 (5353)	0.85 (0.74-0.97)	.02	44.39 (<.001)	73.0%	−2.92 (.003)
Mobile Health System	12 (2662)	0.79 (0.64-0.96)	.02	64.40 (<.001)	82.9%	−2.16 (.03)
Care objective						
Education	10 (5103)	0.86 (0.72-1.02)	.10	39.11 (<.001)	77.0%	−1.82 (.07)
Daily Weight Monitoring	24 (9696)	0.91 (0.83-1.00)	.05	71.77 (<.001)	68.0%	−2.90 (.004)
Monitoring Symptoms	16 (6617)	0.86 (0.74-0.99)	.04	68.81 (<.001)	78.2%	−2.38 (.02)
Medication	15 (4563)	0.83 (0.72-0.95)	.01	47.71 (<.001)	70.7%	−1.55 (.12)
Diet	6 (2569)	0.75 (0.56-1.02)	.07	31.76 (<.001)	84.3%	−1.07 (.29)
Exercise	3 (609)	0.67 (0.35-1.29)	.24	24.38 (<.001)	91.8%	−1.70 (.09)
Depression and Anxiety	2 (464)	0.77 (0.42-1.40)	.39	2.47 (.11)	59.5%	0.00 (<.001)
Care support method						
Call Center Support	23 (9532)	0.91 (0.83-1.00)	.06	71.15 (<.001)	69.1%	−2.83 (.005)
Physician Support	20 (7384)	0.88 (0.78-0.99)	.03	72.83 (<.001)	73.9%	−2.72 (.01)
Automated Alerts	7 (1174)	0.72 (0.53-0.96)	.03	23.59 (<.001)	74.6%	−0.25 (.80)
Collaborative Care Support	6 (2697)	0.89 (0.75-1.07)	.22	12.32 (.03)	59.4%	−0.21 (.83)
Nurse Support	3 (920)	0.80 (0.61-1.03)	.08	3.99 (.13)	49.9%	1.90 (.06)

[a]IRR: incidence rate ratio.

[b]PC: personal computer.

[c]ECG: electrocardiogram.

Table 5. Subgroup meta-analysis to examine the effect of telemonitoring strategies on all-cause hospitalization for randomized controlled trials (RCTs) that did not apply the strategy in the telemonitoring intervention (subgroup 2).

Strategies	RCTs (N participants)	Effect		Heterogeneity		Funnel test Z (P value)
		IRR[a] (95% CI)	P value	Q (P value)	I^2	
Technology application						
PC[b]-based System	23 (8419)	I0.94 (0.86- 1.01)	.13	48.12 (<.001)	54.3%	−2.68 (.007)
Blood Pressure Monitor	6 (2711)	1.00 (0.91- 1.10)	.99	1.78 (.87)	0.0%	−0.53 (.59)
ECG[c] Monitor	18 (7714)	0.90 (0.81- 0.99)	.05	45.19 (<.001)	62.4%	−2.46 (.01)
Heart Rate Monitor	12 (4559)	0.96 (0.85-1.08)	.55	26.57 (<.001)	58.6%	−1.70 (.09)
Mobile Health System	13 (7250)	1.00 (0.94-1.06)	.99	7.90 (.79)	0.0%	−1.04 (.30)
Care objective						
Education	15 (4809)	0.93 (0.84- 1.03)	.21	34.76 (<.001)	59.7%	−2.53 (.01)
Daily Weight Monitoring	1 (216)	N/A[d]	N/A	N/A	N/A	N/A
Monitoring Symptoms	9 (3295)	0.97 (0.91- 1.04)	.44	5.85 (.66)	0.0%	−1.13 (.26)
Medication	10 (5349)	1.02 (0.93-1.12)	.67	15.72 (.07)	42.7%	−2.79 (.01)
Diet	19 (7343)	0.96 (0.88-1.04)	.35	37.33 (<.001)	51.8%	−2.98 (.003)
Exercise	22 (9303)	0.94 (0.86-1.01)	.12	48.04 (<.001)	56.3%	−2.58 (.01)
Depression and Anxiety	23 (9448)	0.91 (0.83-0.99)	.05	71.60 (<.001)	69.3%	−2.94 (.003)
Care support method						
Call Center Support	2 (380)	0.74 (0.47-1.16)	.19	1.59 (.20)	37.2%	0.00 (<.001)
Physician Support	5 (2528)	0.98 (0.90-1.06)	.65	1.78 (.77)	0.0%	−1.05 (.29)
Automated Alerts	18 (8738)	0.98 (0.91-1.05)	.61	33.10 (.01)	48.6%	−2.39 (.02)
Collaborative Care Support	19 (7215)	0.90 (0.81-1.01)	.08	61.24 (<.001)	70.6%	−3.35 (.001)
Nurse Support	22 (8992)	0.92 (0.84- 1.01)	.10	64.19 (<.001)	67.3%	−3.32 (.001)

[a]IRR: risk ratio of mortality.

[b]PC: personal computer.

[c]ECG: electrocardiogram.

[d]N/A: not applicable due to insufficient data.

Table 6. Comparison of the effect of telemonitoring strategies on all-cause hospitalization and all-cause mortality between subgroup 1 and subgroup 2.

Strategies	All-cause hospitalization *P* value	All-cause mortality *P* value
Technology application		
Blood Pressure Monitor	.08	.46
ECG[a] Monitor	.98	.89
Heart Rate Monitor	.19	.92
Mobile Health System	.03	.01
Care objectives		
Education	.45	.92
Monitoring Symptoms	.13	.40
Medication	.02	.59
Diet	.13	.33
Exercise	.33	.28
Depression and Anxiety	N/A[b]	.09
Care support method		
Call Center Support	.37	.73
Physician Support	.14	.35
Automated Alerts	.05	.99
Collaborative Care Support	.92	.28
Nurse Support	.29	.66

[a]ECG: electrocardiogram.

[b]N/A: not applicable due to insufficient data for comparison.

In analysis of all-cause mortality, the mHealth subgroup (subgroup 1, 10 RCTs, n=3711, RR=0.67, 95% CI 0.53-0.85) was also associated with a significantly (*P*=.01) lower RR than the comparison subgroup (subgroup 2, 11 RCTs, n=6852, RR=0.95, 95% CI 0.84-1.07), as shown in Tables 6-8 and Multimedia Appendix 6. Within the mHealth subgroup, the telemonitoring interventions were significantly more effective than usual care on reducing all-cause mortality (subgroup 1, 10 RCTs, n=3711, RR=0.67, 95% CI 0.53-0.85, *P*<.001). No significant heterogeneity was detected in both the mHealth subgroup and comparison subgroup. A significant risk of bias (P_F=.01) was found in the comparison subgroup.

In the subgroup comparison of RR and IRR outcomes, we also used the mixed-effects model to adjust for sex, age, or LVEF, but did not find significant improvements in these RR and IRR analysis results.

Table 7. Subgroup meta-analysis to examine the effect of telemonitoring strategies on mortality in randomized controlled trials (RCTs) that applied the strategy in the telemonitoring intervention (subgroup 1).

Strategies	RCTs (N participants)	Effect		Heterogeneity (P value) and Funnel Test (P value)		Funnel test Z (P value)
		RR[a] (95% CI)	P value	Q (P value)	I^2	
Technology application						
PC[b]-based System	2 (1493)	1.05 (0.90-1.21)	.52	0.52 (.47)	0.0%	0.00 (<.001)
Blood Pressure Monitor	16 (7924)	0.83 (0.71-0.98)	.03	24.22 (.06)	38.1%	−1.03 (.30)
ECG[c] Monitor	7 (3475)	0.82 (0.66-1.01)	.07	7.88 (.24)	23.8%	0.66 (.51)
Telemonitoring Weight Scale	21 (10563)	0.82 (0.71-0.94)	.005	30.89 (.05)	35.3%	−1.89 (.06)
Heart Rate Monitor	11 (6258)	0.82 (0.67-1.00)	.05	19.66 (.03)	49.1%	−1.69 (.09)
Mobile Health System	10 (3711)	0.67 (0.53-0.85)	.001	11.58 (.23)	22.3%	−0.27 (.78)
Care objectives						
Education	9 (6308)	0.81 (0.70-0.93)	.004	7.00 (.53)	0.0%	−1.37 (.17)
Daily Weight Monitoring	21 (10563)	0.82 (0.71-0.94)	.005	30.89 (.05)	35.3%	−1.89 (.06)
Monitoring Symptoms	14 (7640)	0.78 (0.66-0.92)	.004	17.33 (.18)	25.0%	−0.48 (.63)
Medication	12 (5475)	0.77(0.60-0.98)	.04	18.60 (.06)	40.9%	0.12 (.90)
Diet	6 (2569)	0.67 (0.43-1.03)	.07	12.60 (.02)	60.3%	−3.05 (.002)
Exercise	3 (609)	0.56 (0.28-1.13)	.11	2.49 (.28)	19.5%	1.57 (.12)
Depression and Anxiety	2 (464)	0.48 (0.26-0.90)	.02	0.12 (.73)	0.0%	0.00 (<.001)
Care support method						
Call Center Support	19 (10183)	0.81 (0.70-0.93)	.005	28.23 (.05)	36.2%	−2.52 (.01)
Physician Support	17 (8191)	0.78 (0.66-0.92)	.005	28.47 (.02)	43.8%	−2.43 (.01)
Automated Alerts	6 (1102)	0.82 (0.48-1.39)	.46	5.68 (.33)	11.9%	1.67 (.09)
Collaborative Care Support	7 (4235)	0.70 (0.48-1.01)	.06	18.75 (<.001)	68.0%	−1.14 (.26)
Nurse Support	3 (920)	0.69 (0.29-1.67)	.42	5.76 (.05)	65.3%	0.23 (.81)

[a]RR: risk ratio.

[b]PC: personal computer.

[c]ECG: electrocardiogram.

Table 8. Subgroup meta-analysis to examine the effect of telemonitoring strategies on mortality in randomized controlled trials (RCTs) that did not apply the strategy in the telemonitoring intervention (subgroup 2).

Strategies	RCTs (N participants)	Effect		Heterogeneity		Funnel test Z (P value)
		RR[a] (95% CI)	P value	Q (P value)	I^2	
Technology application						
PC[b]-based System	19 (9070)	0.79 (0.68-0.91)	.002	23.55 (.17)	23.6%	–1.08 (.28)
Blood Pressure Monitor	5 (2639)	0.71 0.49-1.03)	.08	6.08 (.19)	34.3%	–1.80 (.07)
ECG[c] Monitor	14 (7088)	0.80 (0.66-0.97)	.03	21.59 (.06)	39.8%	–2.91 (.004)
Heart Rate Monitor	10 (4305)	0.81 (0.65-0.99)	.04	10.50 (.31)	14.3%	–0.59 (.55)
Mobile Health System	11 (6852)	0.95 (0.84-1.07)	.40	11.01 (.35)	9.2%	–1.67 (.09)
Care objectives						
Education	12 (4255)	0.82 (0.65-1.04)	.11	20.66 (.03)	46.8%	–0.89 (.37)
Monitoring Symptoms	7 (2923)	0.89 (0.69-1.13)	.35	8.88 (.18)	32.4%	–2.26 (.02)
Medication	9 (5088)	0.84 (0.70-1.01)	.07	12.11 (.14)	33.9%	–2.58 (.01)
Diet	15 (7994)	0.84 (0.73-0.96)	.02	16.50 (.28)	15.2%	–0.34 (.73)
Exercise	18 (9954)	0.83 (0.72-0.96)	.01	26.24 (.07)	35.2%	–1.46 (.14)
Depression and Anxiety	19 (10099)	0.84 (0.73-0.96)	.02	26.97 (.07)	33.3%	–1.42 (.17)
Care support method						
Call Center Support	2 (380)	1.01 (0.29-3.54)	.98	2.63 (.10)	62.0%	0.00 (<.001)
Physician Support	4 (2372)	0.90 (0.71-1.13)	.37	2.42 (.49)	0.0%	0.34 (.74)
Automated Alerts	15 (9461)	0.81 (0.70-0.94)	.008	25.03 (.03)	44.1%	–3.76 (<.001)
Collaborative Care Support	14 (6328)	0.87 (0.77-0.99)	.04	11.76 (.54)	0.0%	–1.01 (.31)
Nurse Support	18 (9643)	0.85 (0.74-0.96)	.02	23.23 (.14)	26.8%	–2.36 (.02)

[a]RR: relative risk.

[b]PC: personal computer.

[c]ECG: electrocardiogram.

Discussion

Principal Findings

In this systematic review and meta-analysis, we evaluated 18 telemonitoring strategies in 26 RCTs. In addition to a traditional meta-analysis for overall effectiveness, we used a subgroup comparison method to analyze the effects of different telemonitoring components on clinical outcomes. We found that the telemonitoring strategy of providing medication support was associated with reduced all-cause hospitalization, whereas mHealth systems were associated with both reduced all-cause hospitalization and reduced all-cause mortality. Therefore, our review provides unique insight into specific telemonitoring strategies associated with improved clinical outcomes, which will help inform future telemonitoring interventions.

The positive findings related to the medication support strategy underscore the importance of medication therapy in telemonitoring interventions for CHF care. Strong evidence supports the role of modern pharmacological therapy in CHF management for delaying CHF deterioration [66,67], and reducing mortality and hospitalizations [18,67]. However, the

therapeutic benefits are often limited by suboptimal patient adherence [68] and this limitation is not addressed by traditional face-to-face consultations [25]. Our findings suggest that the use of telemonitoring improves the efficacy of medication therapy, possibly through frequent reinforcement of compliance, leading to reduced episodes of clinical deterioration requiring hospitalization. Further research on optimizing medication therapy and underlying care processes in telemonitoring interventions is warranted to improve clinical outcomes in CHF care.

Using the subgroup comparison method, we also found that the strategy of providing telemonitoring interventions through an mHealth system was associated with a significant improvement in both all-cause mortality and hospitalization (or corresponding RR and IRR) outcomes. These positive findings could be supported by several unique advantages of using mHealth for general chronic disease care, including ease of use, portability, and real-time communication [69-71]. These advantages have been shown to improve the underlying care processes of patients' self-management [72], care engagement [73,74], and medication adherence in CHF [75]. Therefore, our positive findings support delivering telemonitoring interventions through

mHealth platforms, consistent with the increasing trend in using smartphones and computer tablets for the primary and secondary prevention of chronic disease [76,77].

Three recent reviews of mHealth in CHF management have resulted in inconsistent outcomes and, consequently, were unable to conclude significant clinical benefits [21,23,78]. In contrast to these traditional reviews, each intervention program in our mHealth subgroup combined both telemonitoring and mHealth interventions. Our positive finding indicates that simple mHealth apps without telemonitoring (enabling care providers to provide timely clinical intervention), such as apps only focusing on self-management or education, were insufficient to improve clinical outcomes. Similarly, this finding suggests that telemonitoring programs focusing on clinical assessment and intervention, but not delivered through an mHealth environment, fail to engage patients with CHF in self-management to the same extent as those provided via mHealth. Therefore, our finding warrants future research on comprehensive care programs combining telemonitoring and mHealth to improve both timely clinical intervention and patient engagement in CHF care.

As a part of our evaluation, we also conducted a traditional meta-analysis to evaluate the overall effectiveness of all of the telemonitoring interventions in the RCTs included in this review. We found that telemonitoring interventions were more effective than usual care on reducing both all-cause mortality and all-cause hospitalizations. This finding adds evidence to support telemonitoring interventions for CHF care generally. In our review, invasive telemonitoring interventions and small RCTs were excluded. These exclusions may have refined the selection of telemonitoring studies, leading to the significant findings, in contrast to the three previous mHealth reviews with inconclusive findings [21,23,78].

It is also important to note that several strategies such as daily weight monitoring, call center support, and exercise contained limited numbers of RCTs in the subgroup or comparison group. The evaluation of these strategies was therefore limited by our subgroup comparison method. However, these strategies should not be overlooked, and further research on their contributions to CHF care, such as improving patient adherence to daily weight monitoring and level of exercise, remains essential to continuously improve telemonitoring outcomes in future studies.

Limitations

Because the objective of our review was to evaluate different telemonitoring strategies, our meta-analysis did not rigorously exclude RCTs with risk of bias, although we did exclude studies with small sample sizes. In addition, this review was an exploratory study, and hence we did not adjust the P value in the multiple comparisons of the telemonitoring strategies.

Conclusions

The issues of mixed mortality and hospitalization outcomes have deterred the adoption of telemonitoring in CHF care. To address this issue, this review extensively investigated strategy-related factors associated with improvements in the outcomes, and found that the strategies of (1) providing medication support and (2) combining telemonitoring interventions through mHealth were associated with a significant improvement in all-cause mortality or hospitalizations. Importantly, these findings emphasize the importance of prioritizing medication therapy and patient engagement through mHealth apps in future telemonitoring interventions for CHF care.

Authors' Contributions

HD, SC, RJ, JD, JL, IY, AM, and IE contributed to conception and design, and in the analysis and interpretation of data. HD, SC, RJ, and AM contributed to drafting of the manuscript and revising it critically for important intellectual content. RJ, JL, AM, and IE contributed to final approval of the manuscript submitted. HD and SC are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Databases and search strategy in the literature search.
[PNG File , 80 KB - jmir_v22i11e20032_app1.png]

Multimedia Appendix 2
The list of excluded studies with the reason for exclusion.
[PNG File , 20 KB - jmir_v22i11e20032_app2.png]

Multimedia Appendix 3
Effectiveness of the strategy of providing medication support on reducing the risk of all-cause hospitalization. The subgroup of randomized controlled trials (RCTs) that provided medication support was compared with the subgroup of RCTs that did not provide medication support.
[PNG File , 469 KB - jmir_v22i11e20032_app3.png]

XSL•FO
RenderX

Multimedia Appendix 4
Effectiveness of the strategy of providing medication support on reducing the risk of all-cause mortality. The subgroup of randomized controlled trials (RCTs) that provided medication support were compared with the subgroup of RCTs that did not provide medication support.
[PNG File , 437 KB - jmir_v22i11e20032_app4.png]

Multimedia Appendix 5
Effectiveness of the strategy of combining with mobile health (mHealth), or applying an mHealth system, on reducing the risk of all-cause hospitalization. The subgroup of randomized controlled trials (RCTs) that applied the mHealth strategy were compared with the subgroup of RCTs that did not apply the strategy.
[PNG File , 322 KB - jmir_v22i11e20032_app5.png]

Multimedia Appendix 6
Effectiveness of the strategy of combining with mobile health (mHealth), or applying an mHealth system, on reducing the risk of all-cause mortality. The subgroup of randomized controlled trials (RCTs) that applied the mHealth strategy were compared with the subgroup of RCTs that did not apply the strategy.
[PNG File , 300 KB - jmir_v22i11e20032_app6.png]

References

1. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. ESC Heart Fail 2014 Sep 30;1(1):4-25. [doi: 10.1002/ehf2.12005] [Medline: 28834669]
2. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]
3. Burnett H, Earley A, Voors AA, Senni M, McMurray JJ, Deschaseaux C, et al. Thirty Years of Evidence on the Efficacy of Drug Treatments for Chronic Heart Failure With Reduced Ejection Fraction: A Network Meta-Analysis. Circ Heart Fail 2017 Jan;10(1):e003529 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.116.003529] [Medline: 28087688]
4. Riley JP, Masters J. Practical multidisciplinary approaches to heart failure management for improved patient outcome. Eur Heart J Suppl 2016 Dec 06;18(suppl G):G43-G52. [doi: 10.1093/eurheartj/suw046]
5. Jeon Y, Kraus SG, Jowsey T, Glasgow NJ. The experience of living with chronic heart failure: a narrative review of qualitative studies. BMC Health Serv Res 2010 Mar 24;10(1):77 [FREE Full text] [doi: 10.1186/1472-6963-10-77] [Medline: 20331904]
6. Desai AS, Stevenson LW. Rehospitalization for Heart Failure. Circulation 2012 Jul 24;126(4):501-506. [doi: 10.1161/circulationaha.112.125435]
7. Butler J, Kalogeropoulos A. Worsening heart failure hospitalization epidemic we do not know how to prevent and we do not know how to treat!. J Am Coll Cardiol 2008 Aug 05;52(6):435-437 [FREE Full text] [doi: 10.1016/j.jacc.2008.04.037] [Medline: 18672163]
8. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. Fam Pract 2017 Apr 01;34(2):161-168 [FREE Full text] [doi: 10.1093/fampra/cmw145] [Medline: 28137979]
9. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
10. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
11. Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 05;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
12. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 09;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
13. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 03;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
14. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of Remote Patient Monitoring After Discharge of Hospitalized Patients With Heart Failure: The Better Effectiveness After Transition -- Heart Failure (BEAT-HF) Randomized Clinical Trial. JAMA Intern Med 2016 Mar;176(3):310-318. [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
15. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011 Jan 9;8(1):30-41 [FREE Full text] [doi: 10.1038/nrcardio.2010.165] [Medline: 21060326]

16. Bui AL, Fonarow GC. Home monitoring for heart failure management. J Am Coll Cardiol 2012 Jan 10;59(2):97-104 [FREE Full text] [doi: 10.1016/j.jacc.2011.09.044] [Medline: 22222071]

17. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland J, Coats A, ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]

18. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Heart 2017 Feb 15;103(4):255-257. [doi: 10.1136/heartjnl-2015-309191] [Medline: 27864319]

19. Chaudhry S, Mattera J, Curtis J, Spertus J, Herrin J, Lin Z. Telemonitoring to improve outcomes after heart failure hospitalization: a randomized controlled trial. Circulation; 2010 Clinical Trial/Clinical Science Abstracts 2010 Nov 23;122(21):2223 [FREE Full text] [doi: 10.1161/CIR.0b013e318200c0b5]

20. Ding H, Jayasena R, Maiorana A, Dowling A, Chen SH, Karunanithi M, et al. Innovative Telemonitoring Enhanced Care Programme for Chronic Heart Failure (ITEC-CHF) to improve guideline compliance and collaborative care: protocol of a multicentre randomised controlled trial. BMJ Open 2017 Oct 08;7(10):e017550 [FREE Full text] [doi: 10.1136/bmjopen-2017-017550] [Medline: 28993389]

21. Cajita MI, Gleason KT, Han H. A Systematic Review of mHealth-Based Heart Failure Interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]

22. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]

23. Carbo A, Gupta M, Tamariz L, Palacio A, Levis S, Nemeth Z, et al. Mobile Technologies for Managing Heart Failure: A Systematic Review and Meta-analysis. Telemed J E Health 2018 Apr 02;24(12):958-968. [doi: 10.1089/tmj.2017.0269] [Medline: 29608430]

24. Conway A, Inglis SC, Clark RA. Effective technologies for noninvasive remote monitoring in heart failure. Telemed J E Health 2014 Jun;20(6):531-538 [FREE Full text] [doi: 10.1089/tmj.2013.0267] [Medline: 24731212]

25. Van Spall HG, Rahman T, Mytton O, Ramasundarahettige C, Ibrahim Q, Kabali C, et al. Comparative effectiveness of transitional care services in patients discharged from the hospital with heart failure: a systematic review and network meta-analysis. Eur J Heart Fail 2017 Nov 24;19(11):1427-1443. [doi: 10.1002/ejhf.765] [Medline: 28233442]

26. Higgins JT, Chandler J, Cumpston M, Li T, Page M. Welch VA Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019).: Cochrane; 2020. URL: www.training.cochrane.org/handbook [accessed 2020-11-06]

27. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, American Heart Association Advocacy Coordinating Committee. Recommendations for the Implementation of Telehealth in Cardiovascular and Stroke Care: A Policy Statement From the American Heart Association. Circulation 2017 Feb 14;135(7):e24-e44. [doi: 10.1161/CIR.0000000000000475] [Medline: 27998940]

28. Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]

29. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care 2013 Jan 09;17(1):R2 [FREE Full text] [doi: 10.1186/cc11919] [Medline: 23302257]

30. IntHout J, Ioannidis JP, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. J Clin Epidemiol 2015 Aug;68(8):860-869 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.03.017] [Medline: 25959635]

31. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013 Apr 24;346(apr24 14):f2304-f2304 [FREE Full text] [doi: 10.1136/bmj.f2304] [Medline: 23616031]

32. Morgan JM, Kitt S, Gill J, McComb JM, Ng GA, Raftery J, et al. Remote management of heart failure using implantable electronic devices. Eur Heart J 2017 Aug 07;38(30):2352-2360 [FREE Full text] [doi: 10.1093/eurheartj/ehx227] [Medline: 28575235]

33. Klersy C, Boriani G, De Silvestri A, Mairesse GH, Braunschweig F, Scotti V, Health Economics Committee of the European Heart Rhythm Association. Effect of telemonitoring of cardiac implantable electronic devices on healthcare utilization: a meta-analysis of randomized controlled trials in patients with heart failure. Eur J Heart Fail 2016 Feb 27;18(2):195-204. [doi: 10.1002/ejhf.470] [Medline: 26817628]

34. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trial 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2]

35. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Stat Med 2010 May 30;29(12):1282-1297. [doi: 10.1002/sim.3602] [Medline: 19408255]

36. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Soft 2010;36(3):1-48.

37. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

XSL•FO
RenderX

38.   Dickersin K, Berlin J. Meta-analysis: state-of-the-science. Epidemiol Rev 1992;14:154-176. [doi: 10.1093/oxfordjournals.epirev.a036084] [Medline: 1289110]

39.   Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343(jul22 1):d4002-d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]

40.   RStudio: Integrated Development for R. 2015. URL: http://www.rstudio.com/ [accessed 2020-11-03]

41.   Yun JE, Park J, Park H, Lee H, Park D. Comparative Effectiveness of Telemonitoring Versus Usual Care for Heart Failure: A Systematic Review and Meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]

42.   Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]

43.   Bekelman DB, Plomondon ME, Carey EP, Sullivan MD, Nelson KM, Hattler B, et al. Primary Results of the Patient-Centered Disease Management (PCDM) for Heart Failure Study: A Randomized Clinical Trial. JAMA Intern Med 2015 May 01;175(5):725-732. [doi: 10.1001/jamainternmed.2015.0315] [Medline: 25822284]

44.   Benatar D, Bondmass M, Ghitelman J, Avitall B. Outcomes of chronic heart failure. Arch Intern Med 2003 Feb 10;163(3):347-352. [doi: 10.1001/archinte.163.3.347] [Medline: 12578516]

45.   Blum K, Gottlieb SS. The effect of a randomized trial of home telemonitoring on medical costs, 30-day readmissions, mortality, and health-related quality of life in a cohort of community-dwelling heart failure patients. J Card Fail 2014 Jul;20(7):513-521. [doi: 10.1016/j.cardfail.2014.04.016] [Medline: 24769270]

46.   Capomolla S, Pinna G, Larovere M, Maestri R, Ceresa M, Ferrari M, et al. Heart failure case disease management program: a pilot study of home telemonitoring versus usual care. Eur Heart J Supp 2004 Nov;6:F91-F98. [doi: 10.1016/j.ehjsup.2004.09.011]

47.   Cleland J, Louis A, Rigby A, Janssens U, Balk A. Noninvasive Home Telemonitoring for Patients With Heart Failure at High Risk of Recurrent Admission and Death: The Trans-European Network–Home-Care Management System (TEN-HMS) Study. ACC Curr J Rev 2005 Sep;14(9):37. [doi: 10.1016/j.accreview.2005.08.225]

48.   Comín-Colet J, Enjuanes C, Verdú-Rotellar JM, Linas A, Ruiz-Rodriguez P, González-Robledo G, et al. Impact on clinical events and healthcare costs of adding telemedicine to multidisciplinary disease management programmes for heart failure: Results of a randomized controlled trial. J Telemed Telecare 2015 Sep 07;22(5):282-295. [doi: 10.1177/1357633x15600583]

49.   Dar O, Riley J, Chapman C, Dubrey SW, Morris S, Rosen SD, et al. A randomized trial of home telemonitoring in a typical elderly heart failure population in North West London: results of the Home-HF study. Eur J Heart Fail 2009 Mar 26;11(3):319-325. [doi: 10.1093/eurjhf/hfn050] [Medline: 19174529]

50.   Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar 18;14(3):333-340. [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]

51.   Giordano A, Scalvini S, Zanelli E, Corrà U, Longobardi GL, Ricci V, et al. Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. Int J Cardiol 2009 Jan 09;131(2):192-199. [doi: 10.1016/j.ijcard.2007.10.027] [Medline: 18222552]

52.   Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL, et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: the Weight Monitoring in Heart Failure (WHARF) trial. Am Heart J 2003 Oct;146(4):705-712. [doi: 10.1016/s0002-8703(03)00393-4]

53.   Hägglund E, Lyngå P, Frie F, Ullman B, Persson H, Melin M, et al. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. Scand Cardiovasc J 2015 Aug 04;49(4):193-199. [doi: 10.3109/14017431.2015.1035319] [Medline: 25968968]

54.   Kalter-Leibovici O, Freimark D, Freedman LS, Kaufman G, Ziv A, Murad H, Israel Heart Failure Disease Management Study (IHF-DMS) investigators. Disease management in the treatment of patients with chronic heart failure who have universal access to health care: a randomized controlled trial. BMC Med 2017 May 01;15(1):90 [FREE Full text] [doi: 10.1186/s12916-017-0855-z] [Medline: 28457231]

55.   Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]

56.   Kotooka N, Kitakaze M, Nagashima K, Asaka M, Kinugasa Y, Nochioka K, HOMES-HF study investigators. The first multicenter, randomized, controlled trial of home telemonitoring for Japanese patients with heart failure: home telemonitoring study for patients with heart failure (HOMES-HF). Heart Vessels 2018 Aug 15;33(8):866-876. [doi: 10.1007/s00380-018-1133-5] [Medline: 29450689]

XSL•FO
RenderX

57. Lyngå P, Persson H, Hägg-Martinell A, Hägglund E, Hagerman I, Langius-Eklöf A, et al. Weight monitoring in patients with severe heart failure (WISH). A randomized controlled trial. Eur J Heart Fail 2012 Apr 18;14(4):438-444. [doi: 10.1093/eurjhf/hfs023] [Medline: 22371525]

58. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a Tablet Computer on Self-care, Quality of Life, and Knowledge: A Randomized Clinical Trial. J Cardiovasc Nurs 2018;33(4):336-343. [doi: 10.1097/JCN.0000000000000462] [Medline: 29369123]

59. Mortara A, Pinna GD, Johnson P, Maestri R, Capomolla S, La Rovere MT, HHH Investigators. Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure). Eur J Heart Fail 2009 Mar 26;11(3):312-318. [doi: 10.1093/eurjhf/hfp022] [Medline: 19228800]

60. Olivari Z, Giacomelli S, Gubian L, Mancin S, Visentin E, Di Francesco V, et al. The effectiveness of remote monitoring of elderly patients after hospitalisation for heart failure: The renewing health European project. Int J Cardiol 2018 Apr 15;257:137-142. [doi: 10.1016/j.ijcard.2017.10.099] [Medline: 29506685]

61. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]

62. Soran OZ, Piña IL, Lamas GA, Kelsey SF, Selzer F, Pilotte J, et al. A randomized clinical trial of the clinical effects of enhanced heart failure monitoring using a computer-based telephonic monitoring system in older minorities and women. J Card Fail 2008 Nov;14(9):711-717. [doi: 10.1016/j.cardfail.2008.06.448] [Medline: 18995174]

63. Villani A, Malfatto G, Compare A, Della Rosa F, Bellardita L, Branzi G, et al. Clinical and psychological telemonitoring and telecare of high risk heart failure patients. J Telemed Telecare 2014 Dec;20(8):468-475. [doi: 10.1177/1357633X14555644] [Medline: 25339632]

64. Wade MJ, Desai AS, Spettell CM, Snyder AD, McGowan-Stackewicz V, Kummer PJ, et al. Telemonitoring with case management for seniors with heart failure. Am J Manag Care 2011 Mar 01;17(3):e71-e79 [FREE Full text] [Medline: 21504262]

65. Wagenaar KP, Broekhuizen BD, Jaarsma T, Kok I, Mosterd A, Willems FF, et al. Effectiveness of the European Society of Cardiology/Heart Failure Association website 'heartfailurematters.org' and an e-health adjusted care pathway in patients with stable heart failure: results of the 'e-Vita HF' randomized controlled trial. Eur J Heart Fail 2019 Feb 28;21(2):238-246. [doi: 10.1002/ejhf.1354] [Medline: 30485612]

66. Sacks CA, Jarcho JA, Curfman GD. Paradigm shifts in heart-failure therapy--a timeline. N Engl J Med 2014 Sep 11;371(11):989-991. [doi: 10.1056/NEJMp1410241] [Medline: 25184412]

67. Ruppar TM, Cooper PS, Mehr DR, Delgado JM, Dunbar-Jacob JM. Medication Adherence Interventions Improve Heart Failure Mortality and Readmission Rates: Systematic Review and Meta-Analysis of Controlled Trials. J Am Heart Assoc 2016 Jun 17;5(6):e002606 [FREE Full text] [doi: 10.1161/JAHA.115.002606] [Medline: 27317347]

68. Zhang Y, Wu S, Fendrick AM, Baicker K. Variation in medication adherence in heart failure. JAMA Intern Med 2013 Mar 25;173(6):468-470 [FREE Full text] [doi: 10.1001/jamainternmed.2013.2509] [Medline: 23400219]

69. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to Use mHealth in Older Adults With Heart Failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]

70. Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]

71. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]

72. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote Monitoring of Patients With Heart Failure: An Overview of Systematic Reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]

73. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov 27;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]

74. Athilingam P, Jenkins B, Johansson M, Labrador M. A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]

75. Piette JD, Striplin D, Marinec N, Chen J, Trivedi RB, Aron DC, et al. A Mobile Health Intervention Supporting Heart Failure Patients and Their Informal Caregivers: A Randomized Comparative Effectiveness Trial. J Med Internet Res 2015 Jun 10;17(6):e142 [FREE Full text] [doi: 10.2196/jmir.4550] [Medline: 26063161]

76. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on EpidemiologyPrevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on CardiovascularStroke Nursing, Council on Functional GenomicsTranslational Biology, Council on Quality of CareOutcomes Research,Stroke Council. Current Science on Consumer Use of Mobile Health for Cardiovascular Disease Prevention: A Scientific Statement From the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]

77.   Ali EE, Chew L, Yap KY. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan
       05;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
78.   Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing
       Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application
       Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline:
       27302310]

Abbreviations

CHF: chronic heart failure
IRR: incidence rate ratio
LVEF: left ventricular ejection fraction
mHealth: mobile health
RCT: randomized controlled trial
RR: risk ratio

XSL•FO
RenderX

Viewpoint

Telemedicine in the United States: An Introduction for Students and Residents

Maryam A Hyder[1]; Junaid Razzak[2], MBBS, PhD

[1]Barnard College, New York, NY, United States

[2]Department of Emergency Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

Corresponding Author:
Junaid Razzak, MBBS, PhD
Department of Emergency Medicine
Johns Hopkins School of Medicine
5801 Smith Ave
Suite 3220
Baltimore, MD, 21209
United States
Phone: 1 410 735 6450
Email: junaid.razzak@jhu.edu

Abstract

Telemedicine refers to the delivery of medical care and provision of general health services from a distance. Telemedicine has been practiced for decades with increasing evidence proving its potential for enhanced quality of care for patients, reduction in hospital readmissions, and increase in savings for both patients and providers. The COVID-19 pandemic has resulted in a significant increase in the reliance on telemedicine and telehealth for provision of health care services. Developments in telemedicine should be structured as complements to current health care procedures, not with the goal of completely digitizing the entire health care system, but rather to use the power of technology to enhance areas that may not be working at their full potential. At the same time, it is also clear that further research is needed on the effectiveness of telemedicine in terms of both financial and patient benefits. We discuss the current and rapidly increasing knowledge about the use of telemedicine in the United States, and identify the gaps in knowledge and opportunities for further research. Beginning with telemedicine's origins in the United States to its widespread use during the COVID-19 pandemic, we highlight recent developments in legislation, accessibility, and acceptance of telemedicine.

Introduction

The World Health Organization (WHO) defines digital health as the use of digital technologies for health purposes, a category that encompasses the increasing use of technologies for health services [1]. Related to digital health is telemedicine; telemedicine seeks to harness the growing role of technology to create more effective health care, and is part of a larger movement of digital health services. Table 1 [1,2] provides definitions of various terms used under the broad scope of digital health and their relationship to each other. These terms have some minor differences, but are important to recognize as relevant to the overall discussion of digital health. Our focus in this paper, however, is specifically telemedicine: its history, its uses, and its importance.

Table 1. Definitions of digital health terms.

Terms	Definitions[a]
Digital health	The use of digital technologies for health. Encompasses eHealth, mobile health (mHealth), and the use of computer science (such as big data and artificial intelligence).
eHealth	The use of information and communication technologies for health.
mHealth	The use of mobile wireless technologies (such as cell phones) for health.
Telemedicine[b]	The use of information and communication technologies to improve patient outcomes by increasing access to care and medical information.
Telehealth	The most basic engagement of eHealth, involving telecommunications and virtual technology to deliver health care outside of traditional facilities.

[a]According to the World Health Organization [1].

[b]The American Telemedicine Association considers telemedicine to be synonymous with telehealth [2].

In the United States, 76% of hospitals connect with patients using some form of telemedicine [3]. Kane and Gillis identified physicians practicing in radiology (39.5%), psychiatry (27.8%), and cardiology (24.1%) as the most frequent users of telemedicine [3].Telemedicine not only is effective for provider-patient interactions but also creates a more connected network between health care professionals. The specialties reported to most commonly use telemedicine to communicate with other health care professionals were emergency medicine (38.8%), pathology (30.4%), and radiology (25.5%) [3]. With increasing need for a multidisciplinary approach to care and patient-provider partnerships, telemedicine has helped further strengthen connections between patients, health care providers, and other stakeholders. In the face of the COVID-19 pandemic, the relevance of telemedicine has become even more acute. Use of telemedicine has been rapidly promoted, and laws on its coverage are rapidly changing.

Here we provide a background of the increasing use of telemedicine in the United States and emphasize its relevance for medical students and residents. Beginning with telemedicine's origins in the United States to its widespread use during the COVID-19 pandemic, we highlight developments in legislation, accessibility, and acceptance of telemedicine. Omitted from this review are comprehensive descriptions of the technical aspects of telemedicine, as well as newer developments in technology such as artificial intelligence. This review is meant to be a starting point for students, and thus we omit more technical information for the sake of brevity and focus.

History of Telemedicine in the United States

In the United States, one of the early uses of telemedicine was established by the National Aeronautics and Space Association (NASA) in 1960, for monitoring astronauts in flight by physicians and medical teams during their mission Project Mercury [4]. NASA designated "medical monitors" to become well-versed in the astronauts' medical history, while conducting research on the effect of the environment of outer space on the human body. Teams of medical observers were positioned at 18 sites across North America, Europe, Africa, and Australia. Their role was to observe and preserve the health of the

astronauts by providing medical advice when needed and consistently evaluating their condition [4]. As NASA demonstrated, using telecommunications to establish contact between health care providers and patients can allow for greater availability of, and access to, health care beyond what was previously conceived as possible. Understanding this great potential that telemedicine holds for increased connectivity, the US National Library of Medicine, in 1966, designated US $42 million for multiple telemedicine projects spanning over 19 years targeted to medically isolated—rural, inner city, and suburban—areas [5].

Approaches to Telemedicine

Telemedicine Use in Various Fields of Medicine

In the United States there are several successful and active models of telemedicine and telehealth, demonstrating their potential. In 2013, there were approximately 4 neurologists per 100,000 people in the country, caring for over 700,000 strokes per year [6]. To deal with the shortage of neurologists, several hospitals and offices implemented telemedicine measures specifically for stroke treatment, or telestroke. Telestroke refers to a common method used in emergency departments to access specialist neurologists, reducing the need for in-house experts. Using telemedicine, neurologists can communicate remotely with emergency physicians and patients with stroke and recommend treatment faster than previously available [6]. This is especially helpful for smaller, rural hospitals that do not have specific specialists for a vast range of conditions compared with larger urban medical centers [7]. Hospitals' use of telestroke enables patients to go to their nearest hospital while still receiving the specialized emergency care they require, and thus all patients are given equal chances of survival [6].

Telemedicine has also been particularly useful in radiology, since nearly all radiology examinations produce digital content, known as teleradiology. In 2014, teleradiology was reported to account for more than half of all telemedicine services performed in the United States [8]. Images and reports collected from either in-person or telemedicine examinations can be transmitted to a remote radiologist, whose report can be sent to the patient's physician or other health care providers. Particularly in areas where there is a shortage of radiologists, teleradiology supplements that shortage by eliminating the

in-person meeting between patient and radiologist, as well as providing quicker readings and results to patients [9].

Even fields that traditionally have depended on face-to-face communications, such as psychiatry, have been able to provide care of comparable quality using telemedicine. As of 2016, a mere 43.1% of the 44.7 million Americans with any sort of mental illness had used mental health services in the United States [10]. Being able to provide care to more Americans with mental illness through telepsychiatry could have a profound effect. The University of Rochester in Rochester, NY, USA created a telepsychiatry program that is available from 8:30 AM to 5:00 PM, Monday through Friday, and is performing about 2000 telepsychiatry consultations per year [10]. A provider simply sends a request for a visit with a patient by computer, and the query is answered by a specialist at the university. The interface of communication is extremely simple, consisting of an iPad on a rolling stand for providers and a video camera for the patient. Nurses are also able to use the interface to record videos of patients in agitated states and securely send the videos to specialists, who can use a patient's video and medical record to recommend treatment and calming techniques [10].

Telemedicine for Patients at Home

The implementation of certain at-home monitoring telemedicine and telehealth systems has also been successful. For example, Columbia University in New York, NY, USA created a home telemedicine unit through its Informatics for Diabetes Education and Telemedicine initiative [11]. This home telemedicine unit is capable of videoconferencing, medical data acquisition, sharing of collected data with physicians, web-based access to clinical data, and web access tailored to diabetes education. Data are transmitted through the existing telephone lines of each household, meaning no extra technological installment was needed. Each home telemedicine unit includes glucose and blood pressure monitors, alerts patients if any of their recorded values are abnormal, and indicates whether they should contact a physician [11]. Long-term outcome data are awaited to determine its full success.

At-home telehealth monitoring, especially for patients with chronic diseases, has proven to be beneficial for not only patients but also hospitals. The Patient Protection and Affordable Care Act requires the US Centers for Medicare & Medicaid Services to penalize readmissions to hospitals that occur within 30 days after discharge [12]. This provision incentivizes hospitals to find another way to communicate with and treat patients who may not require readmission, but simply medical attention. A study conducted by Partners HealthCare in 2014 [13] provided in-home monitoring to over 3000 patients with congestive heart failure. Their personal data—blood pressure, weight, heart rate, pulse—were recorded and uploaded daily to a monitoring system that used decision support software to alert patients and nurses of which patients needed attention. The data were transmitted securely through a telephone service to the internet. Over 6 years, hospital readmissions dropped by 44% for these patients, and the program generated savings of over US $10 million. This program led to a more efficient use of health human resources, as it allowed 3 to 4 nurses to care for 250 patients rather than the 4 to 6 patients they would be assigned to in a traditional

hospital setting [13]. This example demonstrates the potential of telehealth to reduce health care costs, as well as unnecessary hospital admissions for patients, while also benefitting hospitals through reduced readmission penalties. In addition, this study highlights that telemedicine does not benefit one part of the system over another, but ideally works to create overall improvement in the health care system.

Similar results were seen in the implementation of a home health program for veterans, which used telemonitoring to help manage veterans with chronic illnesses, such as diabetes and depression. The Care Coordination/Home Telehealth program yielded high satisfaction from patients and produced a 19% reduction in the number of hospital visits as compared with their usual care [14]. The program generated savings of nearly US $2000 per patient, and even facilitated independent living of 36% of patients, who otherwise would have qualified for long-term residential care [14]. Through this increased access to knowledge, the goal of health care could be shifted from intervention to prevention, as increased education can promote healthier self-management.

Telemedicine as an Emergency Tool During the COVID-19 Pandemic

As COVID-19 continues to spread, many hospitals and physician practices have transitioned to telemedicine to conduct nonessential appointments [15]. A study of trends at NYU Langone Health (New York, NY, USA) showed an 80% decline in in-person visits and a 683% increase in telemedicine visits between March 2 and April 14, 2020 [15]. Most routine consultations could be conducted over the phone with both the patient and the physician at home, which is particularly useful for quarantined physicians. Telemedicine can also play an instrumental role in treating patients in hospitals. For example, Aurora Health, based in Wisconsin, modified emergency department procedures to allow for remote patient intake, which can accommodate faster testing [16]. As the number of COVID-19–infected patients increases, using computers and tablets for telemedicine can also reduce staff exposure in both ambulances and hospitals. Tablets can be effectively disinfected and allow patients to be safely isolated [16]. Electronic intensive care units (e-ICUs) dramatically increase the number of patients that health care workers can monitor at once, and as the numbers of hospitalized patients increase, e-ICU systems can help clinicians manage the load. Finally, developers Avera Health, based in South Dakota, are working on mobile home health care units for sick patients who are able to return home [16]. At-home systems provide patient evaluation without overwhelming an emergency room, allow physicians to assign sick patients to hospital beds through telemedicine, and even potentially facilitate at-home testing [16]. Great advancements in telemedicine are being made to help cope with this global pandemic and demonstrate how the emergence of telemedicine can shift the preparedness infrastructure of the health care system.

Health System Issues for Telemedicine

Reimbursement and the Interest of Health Care Payors and Insurance Providers

The increasing use of telemedicine has led to increasing interest among payors to cover such services. Several insurance companies, and certain government-funded programs, have expanded their policies adjusting for telemedicine services. In 2018, Medicaid in several states broadened the scope of telehealth and telemedicine services for which they reimburse, thus reducing barriers for their use. For example, California approved reimbursement for substance use disorder services delivered through telehealth; Kentucky prohibited the requirement that a physician must be present with the recipient of health care for reimbursement; and Colorado expanded reimbursement to teledentistry [17].

In light of the current opioid epidemic in the United States, telemedicine can provide a newer and quicker avenue for crisis intervention, which is being recognized by payor groups. For example, Connecticut had previously prohibited prescription of controlled substances through telemedicine, but as of 2018 the state has made an exception for the prescription of drugs that treat opioid use disorder or substance use disorder [17]. Additionally, outside of government-funded health care, 31 states and the District of Columbia have enacted parity laws, requiring private insurers to reimburse providers for telehealth services [18]. This shows the impact telemedicine is making on the health care system. Furthermore, legislation currently being proposed is working toward making the practical use of telemedicine and telehealth even easier and more accessible for patients nationwide. Insurance reimbursement poses a large determinant in the acceptance of, and attitude toward, telemedicine. Government insurance reimbursement policies, in particular, have been shown to influence telemedicine adoption.

A survey conducted by the American Telemedicine Association in 2014 found that hospitals were not billing for telemedicine services because government payers would not pay for them, and neither would major private payers [19]. Furthermore, administrative rules differed for in-person versus telemedicine care, another barrier to reimbursement, and insurance companies ultimately followed the guidelines of their individual states, further hindering standardization of telemedicine billing. This study concluded that, if government payers change their policies to include reimbursement for telemedicine services, private payers are also more likely to change their policies [19]. Insurance reimbursement incentivizes telemedicine adoption for hospitals, and thus administrative and state-level rules would likely adjust as well. Finally, billing and coding processes for telemedicine services must also be better understood in order to enact change [19].

State governments have the power to decide whether they will cover telemedicine practices; and states are not required to submit an amendment to Medicaid if they decide to reimburse for telemedicine (in the same way as they do for in-person services). As of 2015, Medicaid reimbursed for telemedicine services in 46 of the 50 states [20]. Medicare, however, has created strict guidelines and instances for which it will reimburse telemedicine practices. To be eligible for Medicare reimbursement for a telemedicine service, a case must conform to a set of national guidelines; namely, the receiver of telemedicine service must be in a rural geographical location defined as either a Health Professional Shortage Area or a county outside of a Metropolitan Statistical Area [20]. These areas are defined by the Health Resources & Services Administration and can be found on their Medicare Telehealth Payment Eligibility Analyzer on their website [21]. As of July 2019, a small number of exceptions to this geographical requirement were enacted, including treating end-stage renal disease, stroke, and substance use disorder [22]. Since Medicare, in particular, serves elderly and disabled populations, telemedicine could alleviate the stress of needing to meet with a physician in person, but the lack of reimbursement deprives these populations of the benefits of telemedicine and telehealth.

In response to COVID-19, Medicare has temporarily expanded its services to cover telemedicine services outside of the prior designated sites and platforms through which one could qualify for reimbursement. As of March 6, 2020, Congress waived the Medicare requirements that limited provision of telehealth care to rural areas [23]. These services can be conducted from any hospital, office, or place of residence—a promising change in policy [24]. Finally, certain Medicare plans cover the cost of COVID-19 testing. These temporary expansions of coverage can, hopefully, become permanent and further promote the use of telemedicine.

Licensing and Use of Telemedicine

In the United States, there is not one standardized license a physician receives that is valid throughout the country, but rather a physician's medical license is issued only for their state of practice. This poses an issue for interstate telemedicine practices, but in 2014 the Federation of State Medical Boards passed the Interstate Medical Licensure Compact to help facilitate interstate practice [20]. The compact allows a physician to apply through their home state medical board for eligibility of a medical license in another state. If the application is approved by their home state, no further verification is required and, after paying a fee to the requested state, they will be issued a medical license [20]. Such legal changes can allow telemedicine to become a "usual" part of the health system in the future. Given the enormity of the COVID-19 health care crisis, many state governments are reducing the licensing restrictions to ensure more effective use of physicians.

Technology and Usability

Concerns about the quality and security of medical records have previously hampered the adoption of telemedicine. In a study conducted by Resneck et al in 2016, medical residents posed as patients and submitted internet pictures to direct-to-consumer telemedicine websites and apps treating skin diseases. In terms of security concerns, they found that none of the websites asked for the patients' ID to prove the photographs were not false [25]. Often clinicians were randomly assigned to patients, depriving patients of any autonomy in their choice of health care provider. In some cases patients were assigned to an international physician who did not have a license to practice

medicine for their state. Even though the services diagnosed the conditions accurately based on images only, diagnoses were incorrect and missed major dermatologic conditions when a more detailed medical history was required for conditions such as secondary syphilis and eczema [25]. Maintaining quality is critical while attempting to expand access to care, and similar standards apply to care provided through telemedicine and to in-person care.

Important aspects of telemedicine and telehealth adoption are the ease of use and lack of technical prerequisites on the part of both patients and health care facilities, as well as reliability of new technologies. If telephone calls are not adequate as a means of making telemedicine visits, access to videoconferencing poses an accessibility barrier both for patients and for health care providers. On the patient side, access to the internet, video cameras, and software poses an issue; and for health care systems, installation of certain videoconferencing software may breach privacy governance policies [25]. In dealing with COVID-19, there is need for immediate conversion to telemedicine, and the time needed to develop new telemedicine platforms that mirror health care workflows is limited. Therefore, videoconference appointments must be monitored and controlled accordingly until access to more appropriate telemedicine software is available [26]. Keeping this in mind, COVID-19 testing must be integrated into telemedicine systems, so that after telemedicine appointments such patients are not immediately sent to the emergency room, further overwhelming hospital systems. For example, testing sites outside of the emergency room should be set up to accommodate telemedicine workflow, whether it be in other office spaces, in tents, or at the patient's home with take-home tests [16].

Finally, despite significant improvement in the reliability and ease of use of technology, there remain a few of limitations, such as some loss of nonverbal communication between physician and patient. This includes, for example, a reassuring touch from a physician or a physician's ability to smell alcohol on a patient's breath and address alcohol consumption during the visit. These limitations could compromise the quality of health care in some cases from both the patient's and physician's point of view [27].

Acceptability by Patients and Providers

For telemedicine to ultimately be successful, it must be accepted by both patients and physicians. In 2016, a survey conducted in Austria asked patients and health professionals to rank what they perceived as the highest- and lowest-ranked benefit of and barrier to eHealth and telemedicine. With a demographic of 43.2% working in health professions and 56.8% in non–health professions, the overall highest-ranked benefit of eHealth was location-independent access to health care services, and the lowest-ranked benefit was better financing of health care. The highest-ranked barrier to eHealth was data security, and the lowest-ranked barrier was increase of administrative burden [28]. The highest-ranked benefit of telemedicine specifically was location-independent access to health care services, and the lowest-ranked benefit was better relationships between doctors and patients. For telemedicine, the highest- and lowest-ranked barriers were the same as those for eHealth [28].

This study also showed that acceptance of telemedicine may vary depending on socioeconomic factors. Participants were categorized by education level, gender, health profession or non–health profession, and digital age group. Digital age group was defined as 2 groups: digital natives, those who grew up in a digital world (defined by age ≤35 years); and digital immigrants, those who did not grow up in a digital world, but adopted the culture [28]. Participants were asked to evaluate their opinion on their knowledge of telemedicine, reliability of health information, reasonability of data exchange, consistent monitoring of patients with chronic disease, and consistent monitoring of all patients. Non–health professionals were 40% less likely to report high knowledge on telemedicine, but having a university degree led to 64% higher odds for high knowledge on telemedicine. The study also found that digital age was a determinant in evaluating the reliability of online health information: digital immigrants were 44% more likely than digital natives to find online information reliable [28]. These findings demonstrated that education on eHealth is crucial for widespread acceptance of eHealth among diverse populations. Since several digital natives were less likely to trust online information or telehealth services as reliable, education is needed even for digital users on not only general definitions of telemedicine and telehealth, but also their practical implementation and reliability.

For telemedicine and telehealth to be implemented successfully, patients must be educated about and comfortable with their use. Systems of informed consent are an essential part of this process, allowing patients to change their consent as they learn more about the technology being used and implemented—especially those with unintended but potential consequences. For example, a study implementing a voice-response telephone counseling system affected some participants in an unanticipated manner. Participants in this study were asked to call a telephone line daily and report their diet and exercise behavior, and most participants spent around 5 minutes per week speaking with the system [25]. Over the course of the study, patients began to form a relationship with the voice, referring to it with gender pronouns, or even naming their relationship as a friend, doctor, or mentor. Some reported the voice of the phone-line as condescending, leaving them with feelings of guilt after their call. After the study ended, some participants reported that they "missed the voice" of the system and continued to call the line [27]. Had patients been informed of the possibility that they could form a relationship—even if subconsciously—with this telephone line prior to the study, they would have been able to make an informed choice of whether to participate. This example helps elucidate the need for education about eHealth not only for medical professionals but also for those who will be interacting with it in order to promote acceptability.

Telemedicine and Telehealth in Special Groups

Patients With Chronic Illnesses

Telemedicine and telehealth can play a large role in care for those with chronic diseases, if approved by both patient and provider. A study conducted in an amyotrophic lateral sclerosis (ALS) clinic held appointments with patients over live videoconferences in place of their regular visit and measured satisfaction surveys from both the patients and care providers. Patients had an overall high level of satisfaction with videoconferencing checkups. They commented that such specialized-care videoconferencing eliminated the need to drive long distances and allowed appointments to be conducted that would have ordinarily been missed or canceled due to bad weather [29]. A symptom of several chronic diseases, including ALS, is fatigue, and thus one of the most popular remarks about the appointments was that videoconferencing was less taxing for patients than their regular in-person visits. Overall, patients felt they received good and high-quality care comparable with an in-person visit. When asked for negative factors of the videoconference appointment, patients stated that they missed the personal physical gestures of reassurance from physicians, such as hugs, during their visits [29].

Elderly Patients

For elderly patients, doctor's appointments often provide an outing; the videoconferencing appointment kept these patients at home, depriving them of an opportunity to mobilize [29]. Suggestions made by the patients included better coordination between clinical care and home care, and thus including a briefing between physician and home caregivers in the videoconference. Additionally, patients commented that the technology and software used may require improvements or replacements, as several patients experienced audio lags during their appointments [29].

Acceptance by Physicians

Geronimo et al [29] also surveyed physicians treating ALS patients for their perspective on the videoconferencing appointments. The most common remark was that patients seemed more comfortable throughout the visit. Additionally, the videoconferencing appointments increased the physicians' understanding of their patients' daily life and the kind of home care they were receiving [29]. The main concern, however, was the lack of physical examination, as well as physical contact

for reassurance. This study indicated that, although telemedicine provides a convenient avenue for health care, it cannot completely replace the current system. Both patients and physicians felt comfortable and satisfied after their visits, but certain crucial elements of an in-person visit are irreplaceable, such as physical examinations and gestures. Acceptance of and attitudes toward this form of telemedicine also varied by geographical location. Rural clinicians especially appreciated the monitoring function of e-ICU systems, as it allowed them to be involved in providing care without being physically present [30].

Acceptance by Nurses

Acceptance depends not only on physicians and patients, but also on the other health care providers interacting with these systems, such as nurses. Health care providers working through various e-ICU systems were surveyed for their general attitudes toward its efficacy and sustainability. The survey found that providers who used the e-ICU more frequently had positive reactions, whereas those who used it less were more likely to have a negative view and to suggest that it imposes a greater burden on nurses [30]. Nurses—who often are responsible for monitoring a large load of patients, as well as coordinating inpatient and outpatients matters—felt the monitoring capabilities of telemedicine helped alleviate some of their stress and to dedicate their time effectively [30]. As this study illustrated, if used effectively, telemedicine can maintain high-quality health care while decentralizing stress on providers.

Future Directions

Table 2 outlines the key future strategies for improving the availability and quality of telemedicine and telehealth services based on a framework from the WHO [1].

Public policy must be compatible with increased eHealth implementation for telemedicine to be practiced regularly. Despite the exceptions to Medicare's telehealth reimbursement policy, the current regulations prevent widespread adoption and use of eHealth and set a precedence for other large private insurers. The Creating Opportunities Now for Necessary and Effective Care Technologies (CONNECT) Act was introduced to the US Senate in 2017, but has yet to make progress. This act proposes the removal of the geographical restrictions under Medicaid and, if passed, would expand coverage of telemedicine services [18]. As a result, telemedicine services would be given the legitimacy of in-person visits in terms of Medicaid and may inspire other insurers to change their private policies.

Table 2. Summary of key strategies for advancement of telemedicine[a].

Domains of World Health Organization framework	Key strategies
Leadership and governance	Advocating for interstate medical licensure
	Promoting a universal reimbursement mechanism for telemedicine in all 50 states for
	Creating public-private telemedicine leadership groups
Strategy and investment	Directing investments for telemedicine programs
	Engaging private health insurance companies
Services and applications	Educating patients
Standards and interoperability	Standardizing informed consent
	Creating a standardized medical licensing system or equivalency system between certain states
Infrastructure	Implementing interoperable systems across health systems
	Developing patient-friendly technologies for home use
Legislation, policy, and compliance	Removing geographic restrictions for Medicaid reimbursement
	Enacting state legislation to require private insurers to reimburse for telemedicine services
Workforce	Training a specific telemedicine health workforce across specialties
	Educating future health professionals in telemedicine

[a]Based on World Health Organization recommendations [1].

State legislation is one of the largest determining factors in telemedicine adoption, and it has been shown that, in states that require private insurers to reimburse for telemedicine, adoption rates have increased [31]. For telemedicine to provide genuine and effective location-independent health care, a standardized licensing system or an equivalency system must also be established in the United States as a whole, or between states. There are several approaches to dealing with interstate licensing other than standardization across the country [20]. State governments can review each other's board examinations and agree to mutually approve medical licensing between them. This solution gives states autonomy in controlling which physicians are eligible to practice telemedicine within their area, while also creating connections between physicians across the country. If legislation is passed legitimizing the role of telemedicine and telehealth in the American health care system, increased education and research on the area can lead to telemedicine being a prominent daily practice for physicians and patients alike.

Conclusions

We have reviewed the definitions, challenges, and potential future directions for telemedicine in the United States. It is no longer a question of whether eHealth has a role to play in health care delivery, rather it is a matter of making it happen. As we are learning during this COVID-19 pandemic, telemedicine and telehealth are critical to ensuring public health and are poised to become reliable and acceptable methods of seeking care for many conditions.

Acknowledgments

Authors' Contributions

MAH conceptualized, researched, drafted, and edited the paper. JR reviewed drafts and helped finalize the paper. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-11-16]
2. American Telemedicine Association. Telehealth: Defining 21st Century Care. Arlington, VA: ATA; 2020. URL: https://www.americantelemed.org/resource/why-telemedicine/ [accessed 2020-11-18]
3. Kane CK, Gillis K. The use of telemedicine by physicians: still the exception rather than the rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]

XSL•FO
RenderX

4.   Link MM. Space Medicine in Project Mercury. NASA SP-4003. NASA Special Publication. Washington, DC: Office of Manned Space Flight, National Aeronautics and Space Administration; 1965. URL: http://articles.adsabs.harvard.edu/full/1965NASSP4003.....L [accessed 2020-11-16]

5.   Li HK. Telemedicine and ophthalmology. Surv Ophthalmol 1999;44(1):61-72. [doi: 10.1016/s0039-6257(99)00059-4] [Medline: 10466589]

6.   Aita MC, Nguyen K, Bacon R, Capuzzi KM. Obstacles and solutions in the implementation of telestroke: billing, licensing, and legislation. Stroke 2013 Dec;44(12):3602-3606. [doi: 10.1161/STROKEAHA.113.001889] [Medline: 24029633]

7.   Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. Int J Med Inform 2015 Sep;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]

8.   Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]

9.   Burute N, Jankharia B. Teleradiology: the Indian perspective. Indian J Radiol Imaging 2009 Feb;19(1):16-18 [FREE Full text] [doi: 10.4103/0971-3026.45337] [Medline: 19774131]

10.   Hasselberg MJ. The digital revolution in behavioral health. J Am Psychiatr Nurses Assoc 2020;26(1):102-111. [doi: 10.1177/1078390319879750] [Medline: 31587624]

11.   Starren J, Hripcsak G, Sengupta S, Abbruscato CR, Knudson PE, Weinstock RS, et al. Columbia University's Informatics for Diabetes Education and Telemedicine (IDEATel) project: technical implementation. J Am Med Inform Assoc 2002;9(1):25-36 [FREE Full text] [doi: 10.1136/jamia.2002.0090025] [Medline: 11751801]

12.   Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]

13.   Kulshreshtha A, Kvedar JC, Goyal A, Halpern EF, Watson AJ. Use of remote monitoring to improve outcomes in patients with heart failure: a pilot trial. Int J Telemed Appl 2010;2010:870959 [FREE Full text] [doi: 10.1155/2010/870959] [Medline: 20508741]

14.   Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]

15.   Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Apr 23;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]

16.   Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Mar 11;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

17.   Dizon R. 2018 in review: state & federal telehealth policy.: National Consortium of Telehealth Resource Centers; 2019. URL: https://www.telehealthresourcecenter.org/2018-in-review-state-federal-telehealth-policy/ [accessed 2020-11-16]

18.   Ladika S. Telehealth overview: the reality check, please. Manag Care 2017 Apr;26(4):16-18 [FREE Full text] [Medline: 28511757]

19.   Antoniotti NM, Drude KP, Rowe N. Private payer telehealth reimbursement in the United States. Telemed J E Health 2014 Jun;20(6):539-543. [doi: 10.1089/tmj.2013.0256] [Medline: 24654748]

20.   Rogove H, Stetina K. Practice challenges of intensive care unit telemedicine. Crit Care Clin 2015 Apr;31(2):319-334. [doi: 10.1016/j.ccc.2014.12.009] [Medline: 25814457]

21.   data.HRSA.gov. Medicare Telehealth Payment Eligibility Analyzer. Rockville, MD: Health Resources & Services Administration URL: https://data.hrsa.gov/ [accessed 2020-11-16]

22.   National policy: Telehealth and Medicare. West Sacramento, CA: Center for Connected Health Policy; 2019. URL: https://www.cchpca.org/telehealth-policy/telehealth-and-medicare [accessed 2020-11-16]

23.   Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020, Public law 116-123-Mar 6, 2020, HR 6074, 116th Congress. 2020. URL: https://www.govinfo.gov/content/pkg/PLAW-116publ123/pdf/PLAW-116publ123.pdf [accessed 2020-11-19]

24.   Medicare.gov. Medicare & coronavirus. Woodlawn, MD: U.S. Centers for Medicare & Medicaid Services URL: https://www.medicare.gov/medicare-coronavirus [accessed 2020-11-16]

25.   Resneck JS, Abrouk M, Steuer M, Tam A, Yen A, Lee I, et al. Choice, transparency, coordination and quality among direct-to-consumer telemedicine websites and apps treating skin disease. JAMA Dermatol 2016 Jul 01;152(7):768-775. [doi: 10.1001/jamadermatol.2016.1774] [Medline: 27180232]

26.   Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]

27.   Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008;17(4):401-416. [doi: 10.1017/S0963180108080535] [Medline: 18724880]

28.   Haluza D, Naszay M, Stockinger A, Jungwirth D. Prevailing opinions on connected health in Austria: results from an online survey. Int J Environ Res Public Health 2016 Aug 11;13(8) [FREE Full text] [doi: 10.3390/ijerph13080813] [Medline: 27529261]

29.   Geronimo A, Wright C, Morris A, Walsh S, Snyder B, Simmons Z. Incorporation of telehealth into a multidisciplinary ALS clinic: feasibility and acceptability. Amyotroph Lateral Scler Frontotemporal Degener 2017 Nov;18(7-8):555-561. [doi: 10.1080/21678421.2017.1338298] [Medline: 28678542]

30.   Ward MM, Ullrich F, Potter AJ, MacKinney AC, Kappel S, Mueller KJ. Factors affecting staff perceptions of tele-ICU service in rural hospitals. Telemed J E Health 2015 Jun;21(6):459-466. [doi: 10.1089/tmj.2014.0137] [Medline: 25734922]

31.   Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]

Abbreviations

ALS: amyotrophic lateral sclerosis
CONNECT: Creating Opportunities Now for Necessary and Effective Care Technologies
e-ICU: electronic intensive care unit
NASA: National Aeronautics and Space Association
WHO: World Health Organization

<u>Original Paper</u>

Investigating Patients' Intention to Continue Using Teleconsultation to Anticipate Postcrisis Momentum: Survey Study

Antoine Grenier Ouimet[1], MSc; Gerit Wagner[1], PhD; Louis Raymond[2], PhD; Guy Pare[1], PhD

[1]Research Chair in Digital Health, HEC Montreal, Montreal, QC, Canada
[2]Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

Corresponding Author:
Guy Pare, PhD
HEC Montreal
3000 Cote-Ste-Catherine Road
Montreal, QC, H3T 2A7
Canada
Phone: 1 5148828672
Email: guy.pare@hec.ca

Abstract

Background: The COVID-19 crisis has drastically changed care delivery with teleconsultation platforms experiencing substantial spikes in demand, helping patients and care providers avoid infections and maintain health care services. Beyond the current pandemic, teleconsultation is considered a significant opportunity to address persistent health system challenges, including accessibility, continuity, and cost of care, while ensuring quality.

Objective: This study aims at identifying the determinants of patients' intention to continue using a teleconsultation platform. It extends prior research on information technology use continuance intention and teleconsultation services.

Methods: Data was collected in November 2018 and May 2019 with Canadian patients who had access to a teleconsultation platform. Measures included patients' intention to continue their use; teleconsultation usefulness; teleconsultation quality; patients' trust toward the digital platform, its provider. and health care professionals; and confirmation of patients' expectations toward teleconsultation. We used structural equation modeling employing the partial least squares component-based technique to test our research model and hypotheses.

Results: We analyzed a sample of 178 participants who had used teleconsultation services. Our findings revealed that confirmation of expectations had the greatest influence on continuance intention (total effects=0.722; $P<.001$), followed by usefulness (total effects=0.587; $P<.001$) and quality (total effects=0.511; $P<.001$). Usefulness (β=.60; $P<.001$) and quality (β=.34; P=.01) had direct effects on the dependent variable. The confirmation of expectations had direct effects both on usefulness (β=.56; $P<.001$) and quality (β=.75; $P<.001$) in addition to having an indirect effect on usefulness (indirect effects=0.282; $P<.001$). Last, quality directly influenced usefulness (β=.34; P=.002) and trust (β=.88; $P<.001$). Trust does not play a role in the context under study.

Conclusions: Teleconsultation is central to care going forward, and it represents a significant lever for an improved, digital delivery of health care in the future. We believe that our findings will help drive long-term teleconsultation adoption and use, including in the aftermath of the current COVID-19 crisis, so that general care improvement and greater preparedness for exceptional situations can be achieved.

Introduction

Background

The COVID-19 pandemic has heightened the urgency of delivering health care through digital technologies [1,2], contributing momentum to existing efforts of transitioning to models of digitally mediated health care provision. As people are encouraged to avoid public places and institutions, telemedicine appears particularly attractive for limiting the propagation of the virus while ensuring the continued care of

XSL•FO
RenderX

patients deemed at risk [1,3-5]. One especially promising telemedicine technology is teleconsultation, which has been considered "capable of reducing emergency room visits, conserving health care resources, and avoiding the spread of COVID-19 by treating patients remotely" [2].

Although the current crisis has induced a high demand, we focus on postpandemic use of teleconsultation and recognize its value for normal conditions. Teleconsultation, or a virtual visit, allows patients to consult with their health care providers online through devices that support text chat and videoconferencing [6]. This type of care delivery is often considered an opportunity to increase accessibility, continuity, and efficiency of care; achieve higher patient satisfaction; and reduce costs [6-9]. At the same time, teleconsultation is potentially comparable, or even preferable, to in-person visits within appropriate medical contexts [9,10].

Prior to the current public health crisis, teleconsultation's general rate of adoption remained low [11-13]. The current situation has caused a profound spike in demand for digitally mediated consultation, and policy makers are breaking down several barriers to foster widespread adoption [14]. As a result, demand for teleconsultation has soared around the world. For instance, the number of teleconsultations in France increased approximately 50-fold in the weeks after it was hit by the pandemic [15].

With the increased demand for teleconsultation observed over the past months, it is an open question whether this momentum will lead to continued adoption in the long term, or whether patients are likely to abandon teleconsultation platforms once the crisis has passed. Previous research has investigated the initial adoption and use of teleconsultation (eg, [12,13,16,17]) but fails to provide explanations of patients' continuance behavior of teleconsultation (ie, the "long-term or sustained use of an IT [information technology] by individual users over a period of time" [18]). Our lack of understanding of the postadoption behaviors of teleconsultation users is troubling since consumers often abandon the use of digital health apps and smart devices shortly after they first use them [19-21]. Given that materialization of the benefits of a technology relies

on its effective and continuous use [22], this dearth of literature is significant in the context of teleconsultation platforms.

Considering these factors, this study aims to answer the following research question: What factors influence a patient's intention to continue using a teleconsultation platform? To address this question, we draw on theories of information systems (IS) adoption and factors specific to teleconsultation services. We empirically tested our research model by surveying employees of two major Canadian companies who have access to the Dialogue teleconsultation platform.

Research Model and Hypotheses Development

To develop our research model (see Figure 1), we draw inspiration from IS research on continuance intention (CI) [23]. We searched the CI literature of the last three years (2015-2017) for peer-reviewed papers [23,24] through three databases: ABI/INFORM, Science Direct, and Web of Science. The search terms were: ("information systems" OR "IS" OR "information technology" OR "IT") AND ("continuance intention"). We identified 221 papers matching our criteria. Specifically, our search of the literature and the descriptive literature review by Nabavi et al [23] resulted in five constructs relevant to the study of teleconsultation continuance by patients: CI, usefulness, quality, trust, and expectations confirmation. In a nutshell, our model contends that users' initial expectations toward an IT are either confirmed or disconfirmed through the user experience [22]. This assessment is captured through the confirmation expectations construct. If the perceived performance of the teleconsultation platform exceeds users' initial expectations, confirmation is said to be positive. In contrast, expectations that are higher than perceived performance produce a negative confirmation. This assessment influences users' expectations toward their future experiences (ie, usefulness, quality, and trust), further affecting their intention to continue using the system [22]. As we aim to explore patients' motivations to continue using teleconsultation, this approach enabled us to focus on the postadoption stage while capturing the preadoption effects of patients' expectations within the use assessment phase of our model.

Figure 1. Research model. H: hypothesis.
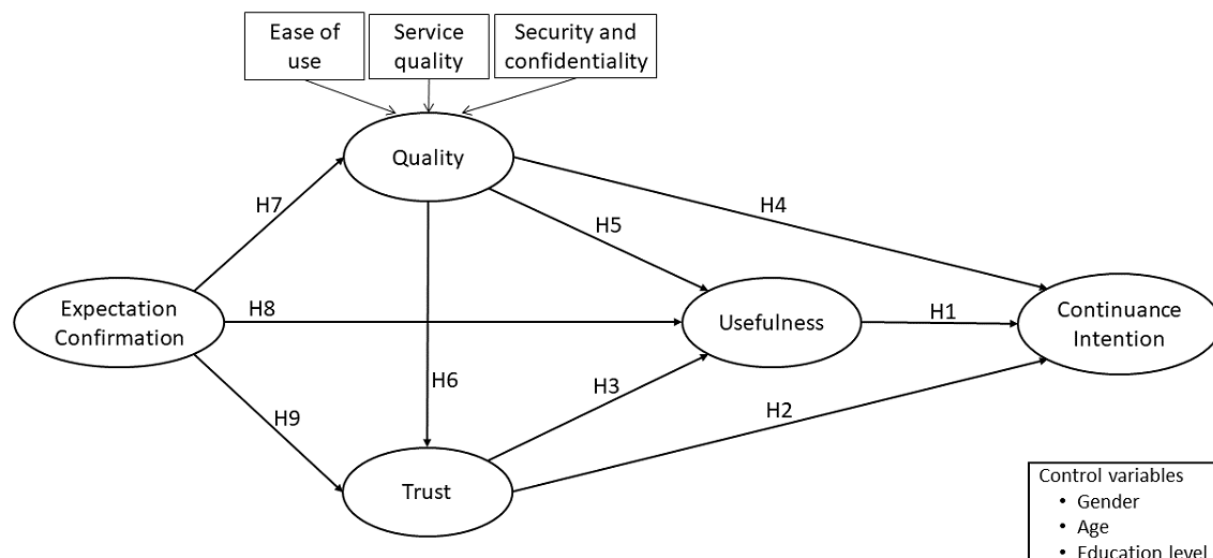
Using CI as a proxy to the actual continuous use behavior is a common practice in IT postadoption research [23]. This is in accordance with research on reasoned and plan behaviors in social psychology [18]. Hence, using CI as our dependent variable is consistent with past work and with our goal of identifying continuous teleconsultation use determinants. In this regard, the perceived usefulness of an IT, described as users' perceptions of the benefits associated with its use [22], has been considered as one of the principal attitudinal beliefs explaining CI [23]. This relationship has been confirmed extensively in the extant literature and in health IT contexts related to teleconsultation such as mobile health (mHealth) [19,25,26]. Considering this, our first hypothesis (H) is as follows:

- H1: Patients' intention to continue using teleconsultation is positively influenced by its usefulness.

IT users' behavioral intentions have also been shown to depend on users' level of trust with a technology and its provider [26-29]. Trust is defined as a party's willingness to be vulnerable to the actions of a second party, whereas the latter is expected to act in the interests of the former, regardless of the former's ability to monitor or control the latter [30]. In other words, trust is the *expectation* that a teleconsultation provider (ie, the Dialogue company), the health professional who assists the patient during the teleconsultation session, and the teleconsultation application can all be counted on. This factor is particularly critical in a virtual environment with high privacy and security risks [28]. These risks are present when users of medical teleconsultation systems disclose information about their physical or mental health, especially since they depend on and have no control over the health care provider [31]. As demonstrated in previous empirical studies, the usefulness of a technology is shaped by other expectations toward it, such as ease of use [32,33]. Therefore, we postulate that trust in a teleconsultation system and its provider affects its usefulness (ie, a teleconsultation service that is deemed untrustworthy is bound to be perceived as having little utility). Hence, we propose the following hypotheses:

- H2: Patients' intention to continue using teleconsultation is positively influenced by their trust in it.
- H3: Usefulness of teleconsultation is positively influenced by patients' trust in it.

Previous empirical evidence indicates that the overall quality of a health care service is of primary importance in shaping the behavioral intent of its users [26,34,35]. Teleconsultation quality is defined here as a patient's judgment of the overall excellence of a teleconsultation. If virtual care is perceived as being of poor quality, it may discourage patients from using it and lead them toward other alternatives such as in-person health services [26]. Much like the assumption that a system that is difficult to use is perceived to be of little usefulness [32], and along with our rationale endorsing the positive influence of trust on usefulness, we postulate usefulness is also affected by the quality of the teleconsultation. In other words, a teleconsultation service perceived as being of poor quality is likely to be regarded as having little utility. In addition, empirical findings demonstrate that users' trust is influenced by the perceived overall quality of the IT service (eg, [26,36]). Hence, we formulate the following hypotheses:

- H4: Patients' intention to continue using teleconsultation is positively influenced by its overall quality.
- H5: Usefulness of teleconsultation is positively influenced by its overall quality.
- H6: Patients' trust in teleconsultation is positively influenced by its overall quality.

Dagger et al [34] found that "customers evaluate service quality at an overall level, a dimensional level, and at sub-dimensional level and that each level drives perceptions at the level above." As a result, the operationalization of this construct should be multidimensional and its measurement formative. Quality has also been theorized as a multidimensional construct in IS research (eg, [37,38]), and we, therefore, look to this field for inspiration to define service quality, system security and confidentiality, and ease of use as first-level dimensions of a teleconsultation quality. Service quality reflects the importance of the service provided to patients [38]. Confidentiality is the

degree to which patients' personal information will not be shared or used against their wishes [39]. Security is defined as patients' perception of another party's ability to protect their private information against loss and unwarranted access [40]. Prior studies have integrated security with constructs like privacy (eg, [41]), and thus, we combined security and confidentiality in a single construct. Ease of use, the degree to which an individual believes that an IT system can be used effortlessly [32], was also incorporated as a formative dimension of quality. Hence, we argue that an increase in the aforementioned dimensions would elevate the perception of the teleconsultation quality for its users and not the opposite. Accordingly, we defined a teleconsultation quality as a second-order construct composed of service quality, system security and confidentiality, and ease of use.

Following findings from the expectation confirmation model (ECM), we included the confirmation of expectations construct in our model. This construct refers to the perceived congruence between the user's initial expectations of the IT use and its actual performance [22]. Its inclusion aims at enabling the comparison of users' initial expectations regarding teleconsultation against the perceived performance resulting from its use. Since usefulness, quality, and trust are expectations toward medical teleconsultation, we posit the following hypotheses:

- H7: Quality of teleconsultation is positively influenced by the confirmation of patients' initial expectations.
- H8: Usefulness of teleconsultation is positively influenced by the confirmation of patients' initial expectations.
- H9: Patients' trust in teleconsultation is positively influenced by the confirmation of their initial expectations.

Given the possibility that various demographic characteristics may influence the attitudes and use behaviors of health IT users (eg, [20,26]), the patients' gender, age, and education level were included as control variables in the research model. Note that socioeconomic variables were not included in the research model given that all employees in the surveyed companies had free access to Dialogue's teleconsultation services as part of their employment insurance benefits. Our research model is presented in Figure 1, and the definitions of the variables included in the model are provided in Multimedia Appendix 1.

Methods

Empirical Setting

The technology studied is the Dialogue teleconsultation platform, which was launched in spring 2016. This teleconsultation service, which is distributed to organizational clients, provides remote health care services to their employees and their employees' relatives (ie, spouse and children). The platform allows patients to communicate with various health care professionals through chat or video calls. Health care professionals can refill prescriptions remotely, refer patients to an external specialist, assist them in navigating the public health care system, and provide medical follow-up from the Dialogue team. During a session, the patient first interacts with an artificial intelligence–based system that carries out an initial

assessment of their needs and redirects the query accordingly. If the system does not refer the patient to another health service, a registered nurse continues the process and discusses the patient's state of health in more depth. This step has three outcomes: immediate diagnosis of the patient, transfer of the patient to a Dialogue physician or nurse for consultation (via chat or video call), or redirection of the patient to a physical examination provided by another health care provider. Ultimately, the patient can either be diagnosed by a Dialogue nurse or physician, or recommended to another health provider that is more suitable to their needs.

The research model was empirically tested with a sample of Dialogue users at two large Canadian companies. The first one is a financial institution with approximately 2000 employees, most of whom are in Quebec. The second company specializes in marketing and advertising, and has close to 500 employees located in different provinces across Canada. At the time of the survey, the 2370 employees (1875 for the first company and 495 for the second) in these two companies had free access to Dialogue teleconsultation services. Both organizations are known for their innovative workplace health management practices, and their staff had access to the teleconsultation platform for over a year before data collection started.

Data Collection

Given our main objectives, we conducted a cross-sectional survey study. The questionnaire was developed and administered using Qualtrics online survey software. The survey was administered in both French and English. Following the design of the questionnaire instrument, a pretest was conducted with 6 users of the teleconsultation system (4 French-speaking and 2 English-speaking users). As a result of the pretest, a few minor changes were made to the questionnaire, and an open-ended question about expectations confirmation was added. All study procedures were approved by the HEC Montreal's research ethics committee.

Operationalization of the Research Variables

The operationalization of the research variables was based on the extant literature. Given the maturity of the research stream, this approach benefits from previous empirical validations and, thus, ensures the validity and reliability of the selected measurement items. Indeed, all the selected items demonstrated appropriate psychometric qualities in prior studies. All items were based on a 5-point Likert scale, ranging from "totally disagree" to "totally agree." Multimedia Appendix 2 presents the items for each construct and their source.

Sample Characteristics

Data were collected between November 1 and November 16, 2018 (first company), and between May 15 and May 28, 2019 (second company). The invitation letters and secure hyperlink to the online questionnaire were posted on the intranet (first company) and sent via email (second company). Reminders to participate in the survey were posted in the middle of the data collection periods. Out of 2370 employees, 354 participated in the survey, representing a response rate of 15%. Of these answers, 44 were excluded due to missing data, resulting in a final sample size of 310 responses. In our sample, 178 (57%)

participants had used the teleconsultation platform at least once. The remainder of this paper is, therefore, based exclusively on their responses, whereas those employees who have not experienced Dialogue (n=132) by the time of the survey could not contribute toward studying continuance.

Results

Profile of the Users and Their Use of the Teleconsultation Platform

The sociodemographic profile of the sampled users is presented in Table 1. Our data set showed that users consulted with Dialogue 3 times on average since registration. Out of the 178 participants, about 43% (n=76) reported using the Dialogue teleconsultation platform three times or more since it became available to them. Such statistics were consistent with the participants' positive health status, as 68% (n=121) of them rated their health as very good or excellent. Frequency analyses and descriptive statistics of the research variables showed a generally positive perception of Dialogue. For instance, user experience was better than expected for 73% (n=130) of patients, and a majority (n=120, 67%) said that using Dialogue provided more benefits than initially expected. In this regard, saving time and gaining better access to health care resources were the main reasons for using Dialogue, as confirmed by 87% (n=154) and 83% (n=147) of patients, respectively. Unsurprisingly, 79% (n=140) of patients reported that Dialogue met their health needs, and 75% (n=133) said they quickly resolved a health concern using teleconsultation. Furthermore, most patients (n=142, 80%) believed Dialogue helped avoid time away from work.

Table 1. Profile of the participants (n=178).

Characteristic	Participants, n (%)
Gender	
Male	47 (26.4)
Female	129 (72.5)
Prefer not to answer	2 (1.1)
Age (years)	
18-24	8 (4.5)
25-34	47 (26.4)
35-44	66 (37.1)
45-54	44 (24.7)
≥55	12 (6.7)
Prefer not to answer	1 (0.5)
Education level	
Secondary school	6 (3.4)
College	40 (22.5)
Certificate	23 (12.9)
Undergraduate level	74 (41.6)
Graduate level	33 (18.5)
Prefer not to answer	2 (1.1)

Assessment of the Measurement Model

We used structural equation modeling (SEM) to validate the research model and test the research hypotheses, employing the partial least squares (PLS) component-based technique for this purpose. PLS-SEM was chosen because of its robustness and lower requirements regarding the distribution of residuals when compared to covariance-based techniques such as EQS and Amos, in addition to being particularly appropriate when the goal is to explain variance [42]. PLS is also able to handle measurement models that include endogenous formative constructs [43].

Table 2 presents the descriptive statistics and reliability coefficients of the research variables. The first step in the SEM analysis is to simultaneously evaluate the measurement model and the structural model with PLS. One must first note in this regard that a single research construct, namely, the quality of the medical teleconsultation, is modeled as being formative due to the composite and multidimensional nature of its conceptualization [44], whereas the other four constructs are reflective [45]. The measurement model also includes another formative construct, patient characteristics, made up of the three control variables, namely, gender, age, and education level.

Table 2. Descriptive statistics and reliability coefficients of the research variables.

Research variable[a]	Items, n	Mean (SD)	Minimum	Maximum	Cronbach alpha[b]	VIF[c,d]
Expectation confirmation	3	3.9 (1.0)	1.0	5.0	.92	N/A[e]
Ease of use	3	4.6 (0.5)	2.3	5.0	.93	1.27
Service quality	3	4.3 (0.8)	1.0	5.0	.92	1.51
Security and confidentiality	4	4.2 (0.7)	2.5	5.0	.95	1.43
Trust	3	4.3 (0.7)	2.1	5.0	.91	N/A
Usefulness	4	4.1 (1.0)	1.0	5.0	.94	N/A
Continuance intention	3	4.4 (0.9)	1.0	5.0	.95	N/A

[a]All variables were measured on 5-point Likert scales (totally disagree to totally agree).

[b]Cronbach alpha coefficient of reliability.

[c]VIF: variance inflation factor.

[d]VIF = $1 / (1 - R_i^2)$, where R_i^2 is the unadjusted R^2 obtained when variable i is regressed against all other variables forming a construct.

[e]N/A: not applicable.

The psychometric properties of the measurements were evaluated in the context of the structural model by assessing the unidimensionality and reliability of the research constructs. First, as presented in Table 2, the Cronbach α values of the research variables varied from .91 to .95, all above the .80 threshold and, thus, confirming their internal consistency. Second, the reflective constructs' item loadings (λ) varied from 0.90 to 0.97, well above the 0.70 threshold and, thus, indicating the unidimensionality of these four constructs. Moreover, the weights (γ) of the three formative variables of the quality construct were all positive and strong enough to be retained in the measurement model [46].

As the usual validity criteria for reflective constructs are inapplicable to a formative construct, one must instead verify that there is no multicollinearity among the formative construct's indicators. One uses the variance inflation factor (VIF) statistic to do so, a common rule being that a variable's VIF value be less than 3.3, or in other words, that less than 70% of the variance in the variable be jointly explained by the other variables [47]. As shown in Table 2, the outer VIF values estimated by PLS for the three formative indicators of the quality construct varied from 1.27 to 1.51, well below the 3.3 threshold. In similar fashion, the three formative indicators of the patient characteristics construct had VIF values varying from 1.01 to 1.13, thus indicating the absence of multicollinearity.

In Table 3, one also finds that the composite reliability coefficient of the reflective constructs varied from 0.94 to 0.97, above the 0.70 threshold and, thus, confirming these constructs' reliability. Their convergent validity was also confirmed, as the average variance extracted (AVE) varied from 0.84 to 0.91, above the 0.50 threshold. The last property to be analyzed in the measurement model, discriminant validity, indicates the extent to which a construct differs from other constructs in the model. In the case of reflective constructs, the shared variance between such a construct and other constructs must be less than the AVE from its indicators, as confirmed in Table 3. In the case of the two formative constructs, quality and patient characteristics, the fact that they shared less than 70% variance with any other related construct in the measurement model, and thus correlated less than perfectly with these constructs, was an indication of such validity [48].

Table 3. Reliability, unidimensionality, and discriminant validity of the research and control constructs.

Construct	CR[a,b]	AVE[c,d]	Items' loading on their respective construct				Interconstruct correlations[e]				
			Item 1	Item 2	Item 3	Item 4	1	2	3	4	5
1. Expectation confirmation	0.95	0.86	0.93	0.93	0.93	N/A[f]	0.93[g]				
2. Quality	—[h]	—	N/A	N/A	N/A	N/A	0.75	N/A			
3. Trust	0.94	0.84	0.92	0.90	0.90	N/A	0.54	0.79	0.92		
4. Usefulness	0.96	0.86	0.92	0.91	0.93	0.95	0.85	0.81	0.63	0.93	
5. Continuance intention	0.97	0.91	0.95	0.94	0.97	N/A	0.70	0.77	0.57	0.82	0.95
6. Patient characteristics	N/A	N/A	N/A	N/A	N/A	N/A	0.11	0.12	0.04	0.11	0.08

[a]CR: composite reliability.

[b]$CR = (\sum \lambda_i)^2 / ((\sum \lambda_i)^2 + \sum(1 - \lambda_i^2))$.

[c]AVE: average variance extracted.

[d]$AVE = \sum \lambda_i^2 / n$.

[e]Subdiagonals: correlation = (shared variance)$^{1/2}$.

[f]N/A: not applicable.

[g]Diagonal: $(AVE)^{1/2} = (\sum \lambda_i^2/n)^{1/2}$.

[h]Inappropriate for formative constructs.

Hypothesis Testing

The research model was tested by evaluating the path coefficients (β) that linked the constructs in the research model (using the SmartPLS software, SmartPLS GmbH), as shown in Figure 2. The model's fit was first assessed by the strength and significance of the path coefficients and the proportion of explained variance in view of PLS's concern with generalization and focus on prediction [49]. The goodness of fit was also assessed by the standardized root mean squared residual index, whose value was 0.06, well below the 0.08 threshold [43].

Figure 2. Research findings.
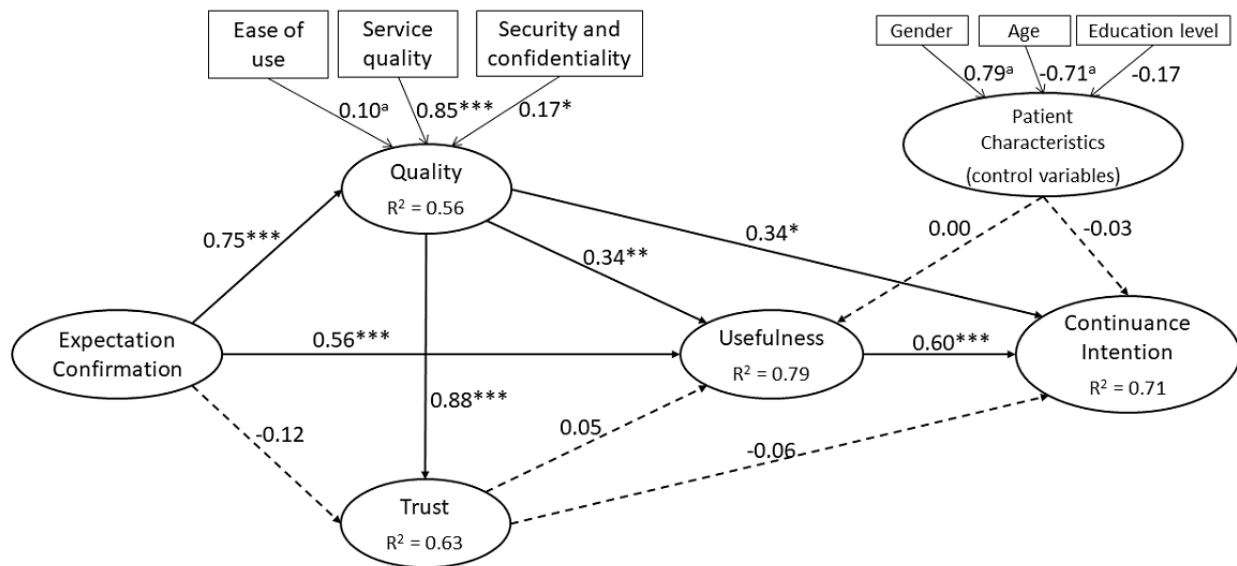
^{a}: $P<.1$ *: $P<.05$ **: $P<.01$ ***: $P<.001$

Hypothesis 1: Confirmed

As seen in Figure 2, a positive and highly significant path coefficient (β=.60, $P<.001$) confirmed that patients who perceive a medical teleconsultation platform to have greater usefulness will have a stronger intent to continue using such a platform. In other words, CI is primarily determined by the benefits patients perceive to have obtained from using teleconsultation.

This initial finding confirms Bhattacherjee's [22] postulate that mHealth IT use is essentially a "means-end" behavior, and this use will continue only if the desired end is achieved.

Hypothesis 2: Unconfirmed

A negative and nonsignificant path coefficient (β=−.06, $P=.27$) disconfirmed the hypothesis that patients who place greater trust in a medical teleconsultation platform will have a stronger intent

to continue using such a platform. It would thus appear that, in the absence of benefits, patients will not necessarily pursue their use of medical teleconsultation technology even though they trust the technology, its provider, and the medical professionals consulted through it.

Hypothesis 3: Unconfirmed

A positive and nonsignificant path coefficient (β=.05, P=.27) did not confirm the hypothesis that patients who place greater trust in a medical teleconsultation platform will perceive such a platform to have greater usefulness. Again, it would appear that the patients' trust in the teleconsultation technology, notwithstanding its presumed criticality in mHealth IT environments, bears no direct consequences upon the personal health benefits obtained from using this technology, nor indirect consequences upon the continuance of such use.

Hypothesis 4: Confirmed

The hypothesis that patients who perceive a medical teleconsultation platform to be of greater quality will have a stronger intent to continue using such a platform was confirmed, as the corresponding path coefficient was shown to be positive and significant (β=.34, P=.01). Looking at the weights of the three components of quality as conceptualized here, this finding highlights the primary importance of the quality of the support services rendered by the platform's provider and the teleconsultation exchanges with the platform's medical professionals (γ=0.85, P<.001). Less important but nonetheless determinant are the patients' perceptions as to the security of the medical teleconsultation platform and the confidentiality of their personal health data (γ=0.17, P=.01), as well as their perceptions of the platform as being easy to learn, use, and master (γ=0.10, P=.07).

Hypothesis 5: Confirmed

A positive and significant path coefficient (β=.34, P=.002) confirmed the hypothesis that patients who perceive a medical teleconsultation platform to be of greater quality will perceive such a platform to have greater usefulness. In line with the ECM [18], this finding demonstrates that the personal benefits obtained from using teleconsultation accrue in part from the quality of the remote health services. The security and confidentiality aspects of the medical teleconsultation technology as well as its ease of use also comes into play, however, less importantly.

Hypothesis 6: Confirmed

A positive and highly significant path coefficient (β=.88, P<.001) confirmed the hypothesis that patients who perceive a medical teleconsultation platform to be of greater quality will place greater trust in such a platform. In line with Akter et al's [26] findings, it appears that it is the quality of the teleconsultation services that are paramount in gaining the patients' trust, with the security and confidentiality of the information transmitted through the teleconsultation being of secondary importance in this regard.

Hypothesis 7: Confirmed

A positive and highly significant path coefficient (β=.75, P<.001) supported the hypothesis that the more the patients' initial expectations related to using a medical teleconsultation platform are confirmed, the better their perception of the quality of such a platform. The patients' perceptions of the platform's quality are deemed to be highly indicative of their affect level [50]. Hence, in line with expectation confirmation theory [18], we found that when confirmation of the patients' initial expectations increases, so does their appreciation of the quality of mHealth IT and, in turn, their appreciation of their IT use experience.

Hypothesis 8: Confirmed

A positive and highly significant path coefficient (β=.56, P<.001) supported the hypothesis that the more the patients' initial expectations related to using a medical teleconsultation platform are confirmed, the better their perception of the usefulness of such a platform. This result is fully in line with a basic proposition of the ECM with regard to the continuance of IT use [22].

Hypothesis 9: Unconfirmed

A nonsignificant and negative path coefficient (β=−.12, P=.15) did not support the hypothesis that the more the patients' initial expectations related to using a medical teleconsultation platform are confirmed, the greater the patients' trust is in such a platform. However, given the preceding results related to hypotheses 7 and 6, it appears that the confirmation of patients' initial expectations has a significant indirect effect upon their trust in the platform, that is, an effect mediated by their perceptions of the platform's quality. Thus, notwithstanding the extent to which their expectations are confirmed, patients would not lose trust in a teleconsultation platform if they perceive the platform to be of enough quality.

Returning to Figure 2, the expectation confirmation and quality constructs were shown to jointly explain more than 70% of the variance in usefulness, a *large* effect size [51]. Similarly, quality and usefulness were shown to jointly explain a large proportion of the variance in CI. Moreover, the three control variables were found to provide no added explanation of the patients' perceptions of the usefulness of the teleconsultation platform nor of their intent to continue using it. Our research model, thus, performed rather well in the nomological integration of the five research constructs. Furthermore, using tests of joint significance of indirect effects [52], we found significant (P<.001) and important (percentage of total effects) indirect effects of expectation confirmation on usefulness, that is, through quality (see Table 4). There were also important indirect effects of quality on patients' CI through usefulness. Finally, we found sizable indirect effects of expectation confirmation on CI, through both quality and usefulness. These last results of the PLS-SEM analysis highlight the "mediating" role played by the patients' assessment of the medical teleconsultation platform's usefulness in explaining their decision to pursue or not pursue their use of such a platform.

Table 4. Breakdown of the total effects of the research constructs.

Relationships between research constructs	Direct effects	Indirect effects	Total effects
Usefulness → continuance intention	0.587	0.000	0.587
Trust → continuance intention	–0.060	0.028	–0.032
Quality → continuance intention	0.339	0.172	0.511
Expectation confirmation → continuance intention	0.000	0.722	0.722
Trust → usefulness	0.048	0.000	0.048
Quality → usefulness	0.340	0.043	0.383
Quality → trust	0.875	0.000	0.875
Expectation confirmation → usefulness	0.571	0.282	0.853
Expectation confirmation → trust	–0.118	0.657	0.539
Expectation confirmation → quality	0.751	0.000	0.751

Discussion

Principal Findings

We examined the influence of usefulness, quality, trust, and expectations confirmation on the CI of a teleconsultation platform in use in Canada. PLS analyses were conducted to test 9 research hypotheses, 6 of which were supported. Based on our findings, the main predictor of teleconsultation CI is perceived usefulness. This is consistent with the extant literature presenting this construct as one of the principal expectations explaining IT continuance [23]. As teleconsultation is exclusively a utilitarian system aiming to provide instrumental value to patients, this finding is also coherent with previous IT use research, which "suggests that utilitarian system-use behaviors are mainly driven by the perceived relevance and usefulness of the system" [53]. Our findings indicate that patients emphasize time savings and quick access to health care professionals as the main benefits. Conversely, the impossibility of remote diagnosis and prolonged waiting times were often presented as the leading causes of low perceived usefulness. In other words, what matters most to patients is that teleconsultation helps them resolve their health issues in an efficient manner.

The inclusion of quality in the research model emerges as an important contribution to the understanding of teleconsultation continuance, considering its significant influence on CI and usefulness. This finding is aligned with extant literature, as several authors observed the influence of quality on users' behavioral intents [26,34,35]. Along with the quantitative results of this study, the qualitative data obtained through an open-ended question indicated that the leading dimension of perceived quality is service quality rather than confidentiality and security, and ease of use. The high level of interactivity and the object of medical teleconsultations cause patients to be concerned with the quality of their interactions and the medical advice received [54]. Hence, service quality, especially in cases involving patients' health status, allows for a more sophisticated understanding of the general concept of quality. The operationalization of quality also benefits from the security and confidentiality construct. As argued by Demiris et al [55], "the ethical issues involved with the application of mobile IT to

health care are many and complex," which makes data security and confidentiality a mandatory requirement for teleconsultation systems. Thus, even if it plays a lesser role than the aspect of service, patients perceive a teleconsultation service to be of high quality if their data is protected from theft, loss, and unauthorized use. Finally, the increased ease people experience when using mobile technologies [19] could explain the relatively low importance of perceived ease of use in building a sense of quality. This construct may prove valuable for the study of populations with lower digital literacies than the one covered by this study.

Confirmation of patients' initial expectations directly influenced teleconsultation usefulness, as expected. Simply put, it is central to determining the usefulness level, especially since it fully accounts for its second drivers (ie, quality) in addition to being its strongest direct predictor. This is captured by the direct and indirect influences of the confirmation of expectations on perceived usefulness reported in Table 4. In addition, expectations confirmation's role in shaping patients' intents should not be underestimated in view of its large indirect effect on CI (Table 4). Indeed, it solely explains the variance in quality and usefulness, which were the only two direct significant predictors of our dependent variable. This is critical, as some users did not seem to fully grasp the possibilities and limitations of the teleconsultation platform, as reflected by many of their comments. For instance, some were surprised at being referred for an in-person examination, given the impossibility of remote diagnosis. In view of the novelty of the service under study and the resulting misunderstanding of its constraints, patients' initial expectations were sometimes unrealistic. Under circumstances where the user experience fell short, usefulness was adversely affected. Therefore, understanding the opportunities and boundaries of online medical consultations is essential for fulfilling expectations, attaining positive levels of usefulness, and sustaining use behavior over time.

Last, in contrast with prior research [26-29], trust was not a significant factor in our model. To explain this unexpected finding, it may be relevant to consider how sensitive the information disclosed by Dialogue users is and how severe the health issues addressed through Dialogue are. Indeed, the data transmitted may not appear sensitive to patients, given the

relatively minor nature of the health issues that can be diagnosed through the type of teleconsultation we studied. Nevertheless, trust may be a significant factor, depending on patients' health issues. For example, it could be relevant in cases of consultation for severe chronic diseases or psychological problems.

Limitations and Suggestions for Future Research

The results of this study should be interpreted with caution, given some methodological limitations. The size of our sample was relatively small. The context-specific nature of this study requires exercising caution about the generalizability of its findings. Organizations providing similar services to their employees may obtain a different picture than the ones presented here. The same applies to teleconsultations offered to patients by hospitals or clinics. Thus, the advancement of knowledge could benefit from similar studies carried out in other contexts. Use context is partially defined by the technology [56], so other teleconsultation technologies are also worth investigating. Should different teleconsultation contexts be studied, we recommend adapting the operationalization of quality (second-order construct) since dimensions of IS success, like system quality, must be anchored in an empirical context [38].

Given the intrinsic limitations of survey research, there may yet remain biases related to the perceptual nature of the research variables' measures. In particular, a measurement model based on a self-administered questionnaire answered by a single respondent may pose a risk of common method bias (CMB) and lead to an overestimation of the relationships between variables [57]. We thus used three post hoc techniques to detect the presence of such bias in our data. First, we used the Harman single factor test [57], that is, we made a principal component factor analysis of all scale items, finding that no single component accounted for 50% or more of the aggregate variance in the research variables and, thus, did not detect any moderate to high levels of CMB. Second, we looked at the correlation matrix between the research and control constructs [58], finding all fifteen correlations to be less than 0.90 and, thus, not showing any major sign of CMB. Third, we used Lindell and Whitney's [59] *marker variable* approach by looking at a variable not assumed to be theoretically linked to any of the research variables. Using the patients' education level for such a purpose, we found the average correlation of this variable with the seven research variables to be –0.065 (with a minimum of –0.09 and a maximum of 0.04) and, thus, not signaling the presence of CMB in our data.

Another limitation of our study was related to its cross-sectional nature, as causality cannot be inferred. We thus encourage future studies on teleconsultation use continuance to employ longitudinal approaches instead. When undertaken in national health care contexts in which teleconsultation services are not free, future studies should also include socioeconomic variables such as patients' income, employment, and place of residence [60]. Finally, our results should be interpreted in view of remote health care contexts that could be altered following the resolution of the COVID-19 crisis.

Contributions of the Study

From a research perspective, our study contributes to the advancement of knowledge by identifying important drivers of patients' intention to continue using teleconsultation (ie, usefulness, quality, trust, and expectations confirmation). We have also demonstrated that it is appropriate to conceptualize quality as a multidimensional second-order construct, reinforcing previous findings (eg, [34]).

Providers of medical teleconsultation and employers interested in them can benefit from this study. This paper identifies the main predictors of continued use for this type of technology, which is crucial to achieving the expected benefits [22]. Given its key role in developing patients' CI, usefulness should be monitored frequently through the measurement of actual or perceived benefits arising from teleconsultation use. In this respect, determinants of usefulness should likewise be monitored. Moreover, as usefulness directly and indirectly relies on confirmation of expectations, patients should be provided with clear information about the possibilities, limitations, and potential benefits of online consultations. Patients surveyed in this study expected to have easier access to physicians via teleconsultation, and a few seemed to be frustrated and disappointed that they could not be remotely diagnosed and were referred to in-person services. The deployment of such a system should include establishing an effective communication plan and involving employers, health care providers, and policy makers, depending on the scale of the deployment.

Teleconsultation providers should also try to reduce waiting times and optimize their customers' health care path since users hope to save time through online medical visits. This is particularly important, as several patients who were instructed to schedule physical examinations felt that the teleconsultation session was a waste of their time. More broadly, our research points to the benefits of focusing on service quality since it is the most important dimension of a teleconsultation's quality. Some users unable to benefit from remote diagnosis suggested that Dialogue could make the appointments they need for them, so the inconveniences associated with an unsuccessful teleconsultation could give way to a simplification of the appointment scheduling process for other health services. The speed and scope of services and the commitment of providers are appealing avenues for sustaining patients' continuous use, regardless of their health status.

Conclusion

This study investigates a technology proving itself to be suitable for public health crisis management [2] as well as being capable of positively addressing challenges of health care systems under normal conditions. We believe that our data set—collected before the COVID-19 pandemic—is more indicative of postcrisis continuance since survey responses collected during the crisis are unlikely to be representative for behavior under normal conditions. We concur that "telemedicine is central to care going forward, not just through this crisis" [14], and is a significant lever for an improved digital delivery of health care in the future. Hence, we hope that our findings will help drive long-term teleconsultation adoption and use, including in the aftermath of the COVID-19 crisis, so that general care

improvement and greater preparedness for exceptional situations can be achieved. The teleconsultation service investigated in this study is particularly relevant in this regard as "[i]ncorporating mHealth in health insurance schemes can help solve the cost and payment barriers and encourage not only clinicians' adoption but also patients' use" [61]. As the current public health crisis has revealed that it is possible to address some of the challenges related to remote health care services, we feel hopeful for the materialization of teleconsultation expected benefits in the long run. Future research efforts should continue focusing on virtual care and teleconsultation systems, and investigate how such technologies contribute to collective well-being.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of constructs.
[DOCX File , 14 KB - jmir_v22i11e22081_app1.docx]

Multimedia Appendix 2
Operationalization of the research variables.
[DOCX File , 16 KB - jmir_v22i11e22081_app2.docx]

References

1. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]

2. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and virtual care trends: cohort study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]

3. Thornton J. Covid-19: A&E visits in England fall by 25% in week after lockdown. BMJ 2020 Apr 06;369:m1401. [doi: 10.1136/bmj.m1401] [Medline: 32253175]

4. Omboni S. Telemedicine during the COVID-19 in Italy: a missed opportunity? Telemed J E Health 2020 Aug;26(8):973-975 [FREE Full text] [doi: 10.1089/tmj.2020.0106] [Medline: 32324109]

5. Magoon V. Operationalizing virtual visits during a public health emergency. Fam Pract Manag 2020;27(3):5-12. [Medline: 32393013]

6. Gordon A, Adamson W, DeVries AR. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. J Med Internet Res 2017 Feb 17;19(2):e35 [FREE Full text] [doi: 10.2196/jmir.6783] [Medline: 28213342]

7. O'Cathail M, Sivanandan M, Diver C, Patel P, Christian J. The use of patient-facing teleconsultations in the National Health Service: scoping review. JMIR Med Inform 2020 Mar 16;8(3):e15380 [FREE Full text] [doi: 10.2196/15380] [Medline: 32175911]

8. Catapan S, Calvo M. Teleconsultation: an integrative review of the doctor-patient interaction mediated by technology. Rev Bras Educ Med 2020;44(1):a. [doi: 10.1590/1981-5271v44.1-20190224.ing]

9. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]

10. Kermani F, Orooji A, Sheikhtaheri A. Teleoncology for children with cancer: a scoping review on applications and outcomes. Int J Med Inform 2020 Jul;139:104118. [doi: 10.1016/j.ijmedinf.2020.104118] [Medline: 32353751]

11. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]

12. Li J, Liu M, Liu X, Ma L. Why and when do patients use e-consultation services? The trust and resource supplementary perspectives. Telemed J E Health 2018 Jan;24(1):77-85. [doi: 10.1089/tmj.2016.0268] [Medline: 28686084]

13. Kane C, Gillis K. The use Of telemedicine by physicians: still the exception rather than the rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]

14. Bashshur R, Doarn C, Frenk J, Kvedar J, Woolliscroft J. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]

15. Croissance record du recours à la téléconsultation en mars. French Health Insurance. 2020. URL: https://www.ameli.fr/espace-presse/communiques-et-dossiers-de-presse/les-derniers-communiques-de-la-caisse-nationale/detail-d-un-communique/3881.php [accessed 2020-07-03]

16. Jiwa M, Meng X. Video consultation use by Australian general practitioners: video vignette study. J Med Internet Res 2013 Jun 19;15(6):e117 [FREE Full text] [doi: 10.2196/jmir.2638] [Medline: 23782753]

17.   Alkmim M, Marcolino M, Figueira R, Sousa L, Nunes MS, Cardoso CS, et al. Factors associated with the use of a teleconsultation system in Brazilian primary care. Telemed J E Health 2015 Jun;21(6):473-483. [doi: 10.1089/tmj.2014.0112] [Medline: 25785650]

18.   Bhattacherjee A, Lin C. A unified model of IT continuance: three complementary perspectives and crossover effects. Eur J Inf Syst 2015 Dec 19;24(4):364-373 [FREE Full text] [doi: 10.1057/ejis.2013.36]

19.   Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]

20.   Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 02;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]

21.   Ledger D, McCaffrey D. Inside wearables part 1: how behavior change unlocks long-term engagement. Endeavour Partners. 2014. URL: https://medium.com/@endeavourprtnrs/ inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2020-07-01]

22.   Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. Manag Inf Syst Q 2001 Sep;25(3):351. [doi: 10.2307/3250921]

23.   Nabavi A, Taghavi-Fard MT, Hanafizadeh P, Taghva MR. Information technology continuance intention: a systematic literature review. Int J E Bus Res 2016;12(1):58-95. [doi: 10.4018/ijebr.2016010104]

24.   Templier M, Paré G. Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. Eur J Inf Syst 2018 Nov 30;27(5):503-550. [doi: 10.1080/0960085x.2017.1398880]

25.   Chen Y, Yang L, Zhang M, Yang J. Central or peripheral? Cognition elaboration cues' effect on users' continuance intention of mobile health applications in the developing markets. Int J Med Inform 2018 Aug;116:33-45. [doi: 10.1016/j.ijmedinf.2018.04.008] [Medline: 29887233]

26.   Akter S, Ray P, D'Ambra J. Continuance of mHealth services at the bottom of the pyramid: the roles of service quality and trust. Electron Markets 2012 May 8;23(1):29-47. [doi: 10.1007/s12525-012-0091-5]

27.   Groß M. Impediments to mobile shopping continued usage intention: a trust-risk-relationship. J Retailing Consumer Services 2016 Nov;33:109-119. [doi: 10.1016/j.jretconser.2016.08.013]

28.   Kim B, Min J. The distinct roles of dedication-based and constraint-based mechanisms in social networking sites. Internet Res 2015 Feb 02;25(1):30-51. [doi: 10.1108/intr-11-2013-0253]

29.   Mou J, Cohen J. Trust and online consumer health service success. Inf Dev 2017 Jul 08;33(2):169-189. [doi: 10.1177/0266666916642507]

30.   Mayer R, Davis J, Schoorman F. An integrative model of organizational trust. Acad Manage Rev 1995 Jul;20(3):709-734. [doi: 10.5465/amr.1995.9508080335]

31.   Gefen D, Straub D. Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. Omega 2004 Dec;32(6):407-424. [doi: 10.1016/j.omega.2004.01.006]

32.   Davis F. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. Manag Inf Syst Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

33.   Hong S, Thong J, Tam K. Understanding continued information technology usage behavior: a comparison of three models in the context of mobile internet. Decision Support Syst 2006 Dec;42(3):1819-1834. [doi: 10.1016/j.dss.2006.03.009]

34.   Dagger T, Sweeney J, Johnson L. A hierarchical model of health service quality. J Serv Res 2007 Jun 29;10(2):123-142. [doi: 10.1177/1094670507309594]

35.   Akter S, D'Ambra J, Ray P. Service quality of mHealth platforms: development and validation of a hierarchical model using PLS. Electronic Markets 2010 Nov 18;20(3-4):209-227. [doi: 10.1007/s12525-010-0043-x]

36.   Ofori K, Boateng H, Okoe F, Gvozdanovic I. Examining customers' continuance intentions towards internet banking usage. Marketing Intelligence Planing 2017 Sep 04;35(6):756-773. [doi: 10.1108/mip-11-2016-0214]

37.   DeLone W, McLean E. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]

38.   DeLone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inf Syst 2003;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]

39.   Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]

40.   Belanche-Gracia D, Casaló-Ariño L, Pérez-Rueda A. Determinants of multi-service smartcard success for smart cities development: a study based on citizens' privacy and security perceptions. Government Inf Q 2015 Apr;32(2):154-163. [doi: 10.1016/j.giq.2014.12.004]

41.   Susanto A, Chang Y, Ha Y. Determinants of continuance intention to use the smartphone banking services. Ind Manage Data Syst 2016 Apr 11;116(3):508-525. [doi: 10.1108/imds-05-2015-0195]

42.   Gefen D, Rigdon EE, Straub D. Editor's comments: an update and extension to SEM guidelines for administrative and social science research. Manag Inf Syst Q 2011;35(2):iii-xiv. [doi: 10.2307/23044042]

43.   Henseler J, Hubona G, Ray PA. Using PLS path modeling in new technology research: updated guidelines. Ind Manage Data Syst 2016 Feb;116(1):2-20. [doi: 10.1108/imds-09-2015-0382]

44. Roberts N, Thatcher J. Conceptualizing and testing formative constructs: tutorial and annotated example. SIGMIS Database 2009 Jul 30;40(3):9-39. [doi: 10.1145/1592401.1592405]

45. Diamantopoulos A, Siguaw JA. Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. Br J Manage 2006 Dec;17(4):263-282. [doi: 10.1111/j.1467-8551.2006.00500.x]

46. Bollen K. Evaluating effect, composite, and causal indicators in structural equation models. Manag Inf Syst Q 2011;35(2):359-372. [doi: 10.2307/23044047]

47. Kock N, Lynn G. Lateral collinearity and misleading results in variance-based SEM: an illustration and recommendations. J Assoc Inf Syst 2012 Jul;13(7):546-580. [doi: 10.17705/1jais.00302]

48. MacKenzie S, Podsakoff P, Jarvis C. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. J Appl Psychol 2005 Jul;90(4):710-730. [doi: 10.1037/0021-9010.90.4.710] [Medline: 16060788]

49. Ringle CM, Sarstedt M, Straub DW. Editor's comments: a critical look at the use of PLS-SEM in "MIS Quarterly". Manag Inf Syst Q 2012;36(1):iii-xiv. [doi: 10.2307/41410402]

50. Forsgren N, Durcikova A, Clay PF, Wang X. The integrated user satisfaction model: assessing information quality and system quality as second-order constructs in system administration. Commun Assoc Inf Syst 2016;38:803-839. [doi: 10.17705/1cais.03839]

51. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hilldale, New Jersey: Lawrence Erlbaum Associates, Inc; 1988.

52. Leth-Steensen C, Gallitto E. Testing mediation in structural equation modeling: the effectiveness of the test of joint significance. Educ Psychol Meas 2016 Apr;76(2):339-351 [FREE Full text] [doi: 10.1177/0013164415593777] [Medline: 29795869]

53. Wu J, Lu X. Effects of extrinsic and intrinsic motivators on using utilitarian, hedonic, and dual-purposed information systems: a meta-analysis. J Assoc Inf Syst 2013 Mar;14(3):153-191. [doi: 10.17705/1jais.00325]

54. Hossain M. Assessing m-Health success in Bangladesh. J Enterprise Inf Manage 2016 Sep 12;29(5):774-796. [doi: 10.1108/jeim-02-2014-0013]

55. Demiris G, Afrin L, Speedie S, Courtney K, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]

56. Hong W, Chan F, Thong J, Chasalow L, Dhillon G. A framework and guidelines for context-specific theorizing in information systems research. Inf Syst Res 2014 Mar;25(1):111-136. [doi: 10.1287/isre.2013.0501]

57. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

58. Pavlou PA, Liang H, Xue Y. Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective. Manag Inf Syst Q 2007;31(1):105. [doi: 10.2307/25148783]

59. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]

60. Arpey NC, Gaglioti AH, Rosenbaum ME. How socioeconomic status affects patient perceptions of health care: a qualitative study. J Prim Care Community Health 2017 Jul;8(3):169-175 [FREE Full text] [doi: 10.1177/2150131917697439] [Medline: 28606031]

61. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR mHealth uHealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]

Abbreviations

AVE: average variance extracted
CI: continuance intention
CMB: common method bias
ECM: expectation confirmation model
H: hypothesis
IS: information systems
IT: information technology
mHealth: mobile health
PLS: partial least squares
SEM: structural equation modeling
VIF: variance inflation factor

<u>Original Paper</u>

Web-Based Data Collection for Older Adults Living With HIV in a Clinical Research Setting: Pilot Observational Study

Katherine Tassiopoulos[1], DSc, MPH[‡]; Carla Roberts-Toler[2], MSc; Carl J Fichtenbaum[3], MD; Susan L Koletar[4], MD

[1]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, United States

[2]Center for Biostatistics in AIDS Research, Harvard TH Chan School of Public Health, Boston, MA, United States

[3]Division of Infectious Diseases, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, United States

[4]Division of Infectious Diseases, Department of Medicine, Ohio State University, Columbus, OH, United States

[‡]for the ACTG A5322 Study Team

Corresponding Author:
Katherine Tassiopoulos, DSc, MPH
Department of Epidemiology
Harvard TH Chan School of Public Health
677 Huntington Avenue
Kresge 817B
Boston, MA, 02115
United States
Phone: 1 617 432 3265
Email: ktassiop@hsph.harvard.edu

Abstract

Background: Longitudinal follow-up of older persons living with HIV is essential for the ascertainment of aging-related clinical and behavioral outcomes, and self-administered questionnaires are necessary for collecting behavioral information in research involving persons living with HIV. Web-based self-reported data collection results in higher data quality than paper-and-pencil questionnaires in a wide range of populations. The option of remote web-based surveys may also increase retention in long-term research studies. However, the acceptability and feasibility of web-based data collection in clinical research involving older persons living with HIV have never been studied.

Objective: This study aims to assess the acceptability and feasibility of a web-based survey to collect information on sexual, substance use, and physical activity behaviors; compare the data quality of the web-based survey with that of a paper-and-pencil questionnaire; and summarize web-based survey metrics.

Methods: This pilot study took place within the AIDS Clinical Trials Group A5322 study, a longitudinal cohort of men and women living with HIV (aged ≥40 years), followed at 32 clinical sites in the United States and Puerto Rico. A total of 4 sites participated in this study. A web-based survey was created using self-administered questionnaires typically completed in A5322 via paper and pencil. Pilot study participants completed these questionnaires via web-based survey at one research visit in lieu of paper-and-pencil administration. Two questions were added to assess feasibility, defined as participants' perception of the ease of web-based survey completion (very hard, hard, easy, very easy), and their preferred format (computer or tablet, paper and pencil, no preference) for completing the questions in the future (acceptability). Feasibility and acceptability were summarized overall and by demographic and clinical characteristics; the proportion of evaluable data by web-based survey versus previously administered paper-and-pencil questionnaires (data quality) was compared for each question.

Results: Acceptability and feasibility were high overall: 50.0% (79/158) preferred computer or tablet, 38.0% (60/158) reported *no preference*, and 12.0% (19/158) preferred paper and pencil; 93.0% (147/158) reported survey completion easy or very easy. Older age was associated with lower odds of preferring computer or tablet to paper and pencil (odds ratio per 1-year increase in age: 0.91, 95% CI 0.85-0.98). Individuals who found the survey hard or very hard had a lower median neurocognitive test score than those who found it easy or very easy. Data quality with web-based survey administration was similar to or higher than that with paper-and-pencil administration for most questions.

Conclusions: Web-based survey administration was acceptable and feasible in this cohort of older adults living with HIV, and data quality was high. Web-based surveys can be a useful tool for valid data collection and can potentially improve retention in long-term follow-up studies.

Introduction

Maintaining sustained research participation is a critical challenge for longitudinal epidemiologic studies [1,2]. Factors that can negatively affect retention include study fatigue, work and personal commitments, and relocation away from the site of a research clinic. These factors may particularly affect studies designed to have extended follow-up durations. For older study participants and those with chronic health conditions, additional factors such as health limitations and difficulty traveling can make attendance at research study visits problematic [3].

Web-based data collection methods that allow the flexibility of remote survey completion may help improve retention in long-term studies [1]. These methods allow for data collection wherever a participant has access to a secure internet connection and ensure confidentiality of responses to potentially sensitive questions, and the embedding of skip patterns and out-of-range response checks and prompts can improve data quality. Data completeness has also been shown to be superior with web-based versus paper-and-pencil data collection [4], and data entry keying errors are reduced because this process is automated [5]. To achieve these data quality advantages, however, surveys must be perceived by study participants as acceptable and easy to complete. This is particularly true with older study participants, who may not be comfortable or familiar with web-based technologies. In 2017, 42% of US individuals aged ≥65 years owned a smartphone, with this proportion decreasing to 31% for those aged 75-79 years and 17% for those aged ≥80 years. In addition, 82% of those aged 65-69 years used the internet, a proportion similar to that of the overall population, but this decreased to 44% for those aged ≥80 years [6]. Particularly among older adults with certain disabilities or activity-limiting impairments, the use of email, texting, and the internet is lower than for those without these impairments [7]. Web-based technologies for research data collection have been used successfully with older adult study populations [8-11], and among older persons living with HIV, web-based and mobile apps have been evaluated for their usefulness and acceptability in health care settings [12-14]. However, to our knowledge, few studies have assessed the feasibility, acceptability, and data quality of web-based technologies for clinical research with older persons living with HIV [15], who may be living with physical and neurocognitive deficits because of their long-term HIV infection [16-20].

The AIDS Clinical Trials Group (ACTG) A5322 study is examining a wide range of clinical and behavioral end points in individuals aging with HIV. Data collection methods include interviews, physical examinations, and chart abstraction as well as self-administered questionnaires that are completed by hand (paper-and-pencil format). Once completed, these questionnaires are handed to research clinic staff in sealed envelopes and mailed to a data management center where responses are keyed. This process can result in incomplete, missed, or lost questionnaires. In addition, although these forms were designed for ease of completion, they do contain free-text responses as well as specific instructions, including those for skip patterns. Participants may have difficulty following the instructions, and the resulting data can be of poor quality because of issues such as skipping questions inappropriately and entering out-of-range free-text responses. Opportunities to clean these data are inherently limited, as data managers cannot query site staff or participants regarding their responses, and as a result, some information in these forms cannot be used.

We piloted a web-based survey that adapted 3 paper-and-pencil questionnaires administered in ACTG A5322. Our overall objective is to determine whether behavioral data collected via paper-and-pencil format could be successfully collected using web-based surveys. This was designed as the initial step of a longer-term goal of incorporating web-based surveys into the study's regular schedule of evaluations, including potential expansion into remote data collection. Our specific aims are to (1) assess the acceptability and feasibility of the web-based survey and identify demographic and health characteristics associated with these measures; (2) compare the data quality of the web-based survey with that of the paper-and-pencil questionnaire; and (3) summarize web-based survey metrics, including frequency of and reasons for survey noncompletion and frequency of *rather not answer* responses.

Methods

Study Population

The A5322 study is an ongoing, long-term observational study following older men and women living with HIV for characterization and evaluation of age-related outcomes. Participants had previously been followed in ACTG A5001, another long-term observational study of participants who had received either their initial HIV antiretroviral treatment medication (treatment naive) or a salvage therapy through an ACTG randomized clinical trial. When the A5001 follow-up ended, participants who had been treatment naive at the time of enrollment in their initial ACTG clinical trial and were aged ≥40 years were eligible to enroll in the A5322 study. Altogether, 1035 participants were enrolled between November 2013 and July 2014 at 32 clinical research sites across the United States, including Puerto Rico. Participants were previously evaluated semiannually (now annually) for immunologic, virologic, and clinical parameters and annually for behavioral parameters. All participants provided written informed consent before enrollment into A5322, and the study was approved by the local institutional review board at each site.

Study Design

A total of 4 sites were chosen for this pilot study based on the number of participants they had enrolled into A5322, the proportion of participants who spoke English as their first language, and the availability of a laptop or desktop with wired

internet and access to a private space in which to complete the survey. All participants at the 4 sites whose primary language was English were eligible to participate in the pilot study. The web-based survey was administered during a single visit, which took place at the 4 sites, in place of the paper-and-pencil questionnaires typically administered.

Survey Development

The survey was developed using Illume, a commercial software tool designed by DatStat (DatStat, Incorporated). The survey was developed only in English and consisted of 3 questionnaires on recent sexual behaviors, current and past substance use, and physical activity (the latter using the International Physical Activity Questionnaire Short Form [21]). At the end of the survey, we included 2 questions to assess acceptability and feasibility (perceived ease of completion). Automatic skip patterns and prompts for participants to re-enter responses that were out-of-range were included. All questions were worded to match the phrasing of the 3 paper-and-pencil questionnaires, except when modifications were necessary to accommodate embedded skip patterns. All questions, except for acceptability and feasibility, included *rather not answer* response options so that participants would not be forced to answer questions and to make preferences for not answering questions transparent.

Outcomes

Frequency of and reasons for noncompletion of the survey were collected on a tracking form completed by the clinic staff. Acceptability was assessed with a three-category variable from the question, "In the future, I would prefer to complete the survey by *computer/tablet, paper/pencil, no preference*." Feasibility was assessed with a four-category variable (*very hard, hard, easy,* and *very easy*).

The data quality of the web-based survey and paper-and-pencil responses was assessed by comparing the proportion of evaluable responses to each question in the web-based survey with that obtained from previously administered paper-and-pencil questionnaires. Evaluable data were defined as any valid response, excluding *rather not answer*. Nonevaluable data were defined as missing, *rather not answer* responses, and out-of-range responses for free-text questions. An example of an out-of-range free-text response would be writing *30* in response to the question, "In the past 7 days, on how many days did you do vigorous exercise?" Logically missing data that resulted from appropriate responses to prior questions were not considered missing (eg, if no vigorous exercise was reported, questions on the number of days and time spent doing vigorous exercise were skipped).

Covariates

Demographic and functional fitness characteristics were assessed at the visit closest to the survey administration unless otherwise indicated. The variables included age, race/ethnicity, sex, education (assessed during A5001 follow-up), history of comorbidities (diabetes, kidney disease, liver disease, cardiovascular disease, stroke, hepatitis C–positive serology, and cancer [within 5 years]), frailty, disability; and neurocognitive function and impairment. Frailty was assessed using the Fried Frailty assessment, which includes 4-m walk

speed; grip strength; and self-reported unintentional weight loss, exhaustion, and low activity [22]. Individuals meeting 3 to 5 components are categorized as frail, those meeting 1 to 2 components are categorized as prefrail, and those meeting 0 components are categorized as nonfrail. For this analysis, we used a two-category frailty variable (frail vs nonfrail/prefrail). Disability was assessed with the Lawton-Brody Instrumental Activities of Daily Living (IADL) questionnaire using self-reported limitations in performing 8 tasks: housekeeping, money management, cooking, transportation, telephone use, shopping, laundry, and medication management [23]. We defined IADL disability as ≥1 limitation. We assessed neurocognitive function using the Trail-Making Tests A and B and the Wechsler Adult Intelligence Scale-Revised Digit Symbol subtest. The raw scores from these 3 evaluations were standardized by age, sex, race/ethnicity, and education and combined into one summary z-score (NPZ-3 score). Neurocognitive impairment was defined as having at least one z-score ≥2 SDs below the mean or at least two z-scores ≥1 SD below the mean. Raw scores were normalized only for participants who were Black, White, or Hispanic.

Statistical Analysis

Survey Metrics

To compare the frequency of *rather not answer* responses by acceptability and feasibility, the total number of *rather not answer* responses given by each participant across all survey questions was summed, and this summed value was then dichotomized to 0 or ≥1 *rather not answer* responses. Chi-square tests were used for comparisons.

Acceptability and Feasibility

We compared the distribution of acceptability by demographic and health characteristics using chi-square tests for categorical variables and the Kruskal-Wallis test for continuous variables. Multinomial logistic regression models were fit to evaluate the association of these variables with acceptability as a three-category variable, *paper and pencil* as the reference category in one model, and *no preference* as the reference in the second model. Univariate models were first fit with each individual variable. All variables with a *P* value of less than .10 in the univariable model were then included in a multivariable model. Feasibility was compared by age and neurocognitive score using the Wilcoxon test. As only a few individuals reported that the survey was hard or very hard, feasibility was evaluated as a dichotomous variable (easy/very easy vs hard/very hard).

Data Quality

We compared the proportion of evaluable responses to each question by survey format using chi-square tests. The first comparison was between pilot study participants' web-based responses and their responses to the most recently completed paper-and-pencil questionnaire. A second comparison was between pilot study participants' web-based responses and responses of all A5322 participants on their most recently completed paper-and-pencil questionnaire. Chi-square tests were used for these comparisons instead of a matched approach because the number of responses to many of the questions

differed by mode of administration. Therefore, it was not possible to use a matched approach that would take into account within-person correlations.

Finally, we compared the proportion of evaluable data for all questions within the web-based survey by age and neurocognitive impairment.

SAS 9.4 was used for all analyses (SAS Institute).

Results

Survey Metrics

A total of 180 participants at the 4 sites were eligible to participate in the pilot study. Of these 180 participants, 159 (88.3%) completed the web-based survey; for the 21 eligible participants who did not complete the survey, the following reasons were provided: participant declined (8/180, 4.4%), clinic error (7/180, 3.9%), technical difficulties (3/180, 1.7%), time constraints (2/180, 1.1%), and cognitive impairment (1/180, 0.6%).

Of the 35 questions in the web-based survey that included a *rather not answer* response option, 18 questions (51%) received at least one (range 1-4) *rather not answer* response. Multimedia Appendix 1 lists the questions that had one or more *rather not answer* responses. Of 159 participants, 24 (15.1%) responded *rather not answer* to one or more of the survey questions, with 16 responding *rather not answer* to 1 question, 6 responding to 2 questions, and 2 responding to 3 questions.

Acceptability and Feasibility

Table 1 summarizes the participant characteristics, overall and by acceptability. Most participants were men, and the median age was 54 years (IQR 49-61). Most participants had more than a high school education, and the majority reported no IADL limitations.

Table 1. Demographic and functional fitness by acceptability.

Characteristic	Prefer computer/tablet (n=79)	Prefer paper-and-pencil questionnaire (n=19)	No preference (n=60)	Total (N=158[a])	P value
Age (years), median (Q1-Q3)	53 (48 to 60)	59 (52 to 66)	55 (50 to 61.5)	54 (49 to 61)	.04[b]
Sex, male[c], n (%)	62 (79)	14 (74)	53 (88)	129 (81.7)	.21[d]
Race and ethnicity, n (%)					.75[d]
White, non-Hispanic	43 (54)	13 (68)	39 (65)	95 (60.1)	
Black, non-Hispanic	30 (38)	6 (32)	19 (32)	55 (34.8)	
Hispanic (regardless of race)	3 (4)	0 (0)	2 (3)	5 (3.2)	
Asian, Pacific Islander	2 (3)	0 (0)	0 (0)	2 (1.3)	
More than one race	1 (1)	0 (0)	0 (0)	1 (0.6)	
NPZ-3 score: median (Q1-Q3)[e]	0.55 (−0.20 to 1.20)	0.50 (−0.20 to 1.20)	0.75 (0.10 to 1.25)	0.60 (−0.10 to 0.20)	.34[b]
Neurocognitive impairment[e,f], n (%)	9 (11)	2 (11)	4 (7)	15 (9.5)	.59[d]
Greater than high school education level[f], n (%)	48 (61)	12 (63)	43 (72)	103 (65.2)	.55[d]
Frail, n (%)	8 (10)	1 (5)	5 (8)	14 (8.9)	.79[d]
≥1 IADL[g] limitation[f], n (%)	19 (24)	2 (11)	4 (7)	25 (15.8)	.02[d]
History of any comorbidity, n (%)	68 (86)	16 (84)	52 (87)	136 (86.1)	.96[d]

[a]One participant (Black non-Hispanic male, high school education, no IADL limitations) skipped the acceptability question.

[b]Kruskal-Wallis test.

[c]Sex at birth. Information on gender not available.

[d]Chi-square test.

[e]Black, White, and Hispanic participants only.

[f]The following variables had missing observations: neurocognitive impairment (N=3, all chose computer/tablet), education (N=3, all chose computer/tablet), and IADL limitations (N=1, chose *paper and pencil*, and N=2, *no preference*).

[g]IADL: Instrumental Activities of Daily Living.

Overall, 50.0% (79/158) of participants indicated that they would in the future prefer to answer the questionnaires via computer or tablet, 12.0% (19/158) said they would prefer the *paper-and-pencil* format, and 38.0% (60/158) had *no preference*. The median (IQR) age of participants who preferred the *paper-and-pencil* questionnaire was 59 years (IQR 52-56), compared with 53 years (IQR 48-60) for those who preferred computer or tablet and 55 years (IQR 50-61.5) for those who indicated *no preference*. Those with ≥1 IADL limitation were more likely to prefer the computer or tablet option (19/25, 76%)

than either *paper and pencil* (2/25, 8%) or *no preference* (4/25, 16%), whereas those with no self-reported limitations were similarly as likely to prefer computer or tablet (60/131, 45.8%) or have *no preference* (55/131, 42.0%). There were no differences in acceptability by other characteristics.

In an adjusted multinomial logistic regression model including age and IADL, with *paper and pencil* as the reference group,

older age was associated with lower odds of preferring the computer or tablet format (odds ratio [OR] per 1-year increase in age 0.91, 95% CI 0.85-0.98; Table 2). With *no preference* as the reference group, those with ≥1 IADL were more likely to prefer the computer or tablet (OR 4.42, 95% CI 1.40-13.9; Table 3).

Table 2. Multinomial logistic regression model: factors associated with acceptability (paper and pencil as the reference group).

Variable and outcome	OR[a] (95% CI)	*P* value
Age at survey completion, years		
Computer/tablet	0.91 (0.85-0.98)	.01
No preference	0.95 (0.89-1.02)	.14
IADL[b] impairment (≥1 impaired categories)		
Computer/tablet	2.72 (0.55-13.5)	.22
No preference	0.62 (0.1-3.74)	.60

[a]OR: odds ratio.
[b]IADL: Instrumental Activities of Daily Living.

Table 3. Multinomial logistic regression model: factors associated with acceptability (no preference as the reference group).

Variable and outcome	OR[a] (95% CI)	*P* value
Age at survey completion, years		
Computer/tablet	0.96 (0.92-1.01)	.11
Paper-and-pencil questionnaire	1.05 (0.98-1.13)	.14
IADL[b] impairment (≥1 impaired categories)		
Computer/tablet	4.42 (1.4-13.9)	.01
Paper-and-pencil questionnaire	1.62 (0.27-9.87)	.60

[a]OR: odds ratio.
[b]IADL: Instrumental Activities of Daily Living.

Most participants (147/158, 93.0%) reported that the web-based survey was easy or very easy to complete. Individuals who found the survey hard or very hard had a lower median NPZ-3 score than did those who found it easy or very easy; age was not associated with feasibility (Table 4).

Table 4. Age and NPZ-3 score by feasibility.

Characteristics	Hard/very hard (n=10)	Easy/very easy (n=148)	Total (N=158[a])	*P* value[b]
NPZ-3 score				
Minimum to maximum	−1.10 to 3.00	−2.00 to 3.20	−2.00 to 3.20	
Median (Q1-Q3)	−0.30 (−1.00 to 0.20)	0.70 (0.00 to 1.20)	0.60 (−0.10 to 1.20)	.02
Age at survey completion				
Minimum to maximum	44 to 77	44 to 79	44 to 79	
Median (Q1 to Q3)	49.5 (48.0 to 52.0)	54.5 (49.5 to 61.0)	54.0 (49.0 to 61.0)	.16

[a]One participant skipped the feasibility question.
[b]Wilcoxon test.

Participants who thought the web-based survey was easy or very easy to complete were more likely to prefer to answer future questions via computer or tablet than those who found it hard or very hard (*data not shown*). There were no differences

in acceptability or feasibility between participants with ≥1 versus no *rather not answer* responses (Multimedia Appendix 2).

XSL•FO
RenderX

Data Quality

Comparisons of evaluable responses by questionnaire administration are summarized in Table 5. Although overall data quality was high for both formats, for most questions, the web-based survey yielded a proportion of evaluable data that was similar to or greater than for the paper-and-pencil questionnaire. Data quality was higher with the web-based survey for most of the sexual behavior questions and all the physical activity questions; the greatest difference in the latter was for questions about the length of time performing various physical activities. The only 2 questions for which data quality was low in both formats were anal sex with women with condom

and oral sex with women with condom. These were endorsed by a few individuals (N=14, web-based survey; N=8, *paper-and-pencil* questionnaire) such that the number of missed responses resulted in very low proportions of evaluable responses. Almost all substance use behavior responses were similarly evaluable in both formats. The results were similar when comparing the responses of those who completed the web-based survey with the paper-and-pencil responses of all A5322 participants (Multimedia Appendix 3). We found no differences in data quality when we examined the proportion of evaluable data *within* the web-based survey by age and neurocognitive impairment (*data not shown*).

Table 5. Evaluable responses by questionnaire format (N=159).

Question	Paper and pencil: proportion of evaluable responses		Web-based survey: proportion of evaluable responses		*P* value
	n (%)	N	n (%)	N	
Sexual behavior					
Any sexual partners	140 (88.1)	159	155 (97.5)	159	.001
Number of sexual partners	67 (73)	92	75 (90)	83	.003
Oral sex with a man	81 (88)	92	65 (71)	92	.004
Oral sex with a man with condom	59 (82)	72	57 (90)	63	.15
Oral sex with a woman	73 (79)	92	77 (93)	83	.01
Oral sex with a woman with condom	8 (28)	29	7 (54)	13	.10
Vaginal sex	75 (82)	92	78 (94)	83	.01
Vaginal sex with condom	22 (54)	41	23 (79)	29	.03
Anal sex with a man	76 (83)	92	79 (95)	83	.009
Anal sex with a man with condom	40 (69)	58	41 (91)	45	.007
Anal sex with a woman	69 (75)	92	79 (95)	83	<.001
Anal sex with a woman with condom	1 (4)	24	3 (43)	7	.007
Any new sexual partners	83 (90)	92	79 (95)	83	.21
Number of partners who know your HIV status	84 (91)	92	75 (90)	83	.83
Number of partners with known HIV status	86 (93)	92	77 (93)	83	.85
Substance use					
How often drink alcohol	159 (100.0)	159	158 (99.4)	159	.32
How many drinks containing alcohol	99 (98.0)	101	96 (96.0)	100	.40
How often binge drink alcohol	101 (100.0)	101	99 (99.0)	100	.31
Not getting things done because of alcohol	98 (97.0)	101	99 (99.0)	100	.32
Emotional problems from alcohol	96 (95.0)	101	100 (100.0)	100	.02
Last time used tobacco	156 (98.1)	159	157 (98.7)	159	.65
Last time used marijuana	157 (98.7)	159	158 (99.4)	159	.56
Last time used cocaine	157 (98.7)	159	158 (99.4)	159	.56
Last time used heroin	156 (98.1)	159	159 (100.0)	159	.08
Last time used amphetamines	154 (96.9)	159	159 (100.0)	159	.02
Last time used other nonprescribed substance	152 (95.6)	159	159 (100.0)	159	.007
Not getting things done because of substance use	86 (97)	89	98 (98.0)	100	.56
Emotional problems from substance use	85 (96)	89	98 (98.0)	100	.33
Physical activity					
How many days spent doing vigorous activities	153 (96.2)	159	159 (100.0)	159	.01
How much time spent doing vigorous activities	79 (90)	88	85 (98)	87	.03
How many days spent doing moderate activities	149 (93.7)	159	159 (100.0)	159	.001
How much time spent doing moderate activities	91 (82.7)	110	102 (99.0)	103	<.001
How many days spent walking ≥10 min	148 (93.1)	159	159 (100.0)	159	<.001
How much time spent walking	119 (87.5)	136	140 (99.3)	141	<.001
How much time spent sitting	126 (79.2)	159	156 (98.1)	159	<.001

XSL•FO
RenderX

Discussion

Principal Findings

To our knowledge, this is one of the first studies on the feasibility and acceptability of web-based data collection for clinical research purposes among older adults living with HIV, and our findings indicate that web-based surveys can successfully be implemented in research with this population. The perceived acceptability and feasibility of the web-based survey were high, and almost all participants found the survey to be easy or very easy to complete. The data quality of responses via the web-based survey was similar to or higher than that for the same questions in the paper-and-pencil questionnaires.

Although overall acceptability and feasibility were high, there were differences by demographic and clinical factors that need to be taken into account when deciding how to incorporate web-based data collection. Younger adults were more likely than older adults to prefer a web-based format. Although the adoption of web-based technologies has steadily increased in recent years among older adults in the United States, it continues to lag behind those of younger adults, as does confidence in one's ability to use these technologies [6]. Although age was not associated with the feasibility of web-based survey completion, neurocognitive function was; participants with lower neurocognitive test scores were less likely to report that the survey was easy or very easy to complete.

Although data quality within the web-based format was not affected by either age or neurocognitive function, the fact that participant preference did depend on these factors has to be considered when determining how to incorporate web-based surveys in a clinical research study. A mixed mode study design is likely needed, with participants given the option of either web-based or paper-and-pencil administration [24]. This option will become increasingly important, as participants continue to age and perhaps experience impairments in neurocognitive function. A growing body of literature indicates the equivalence of paper and pencil and electronic methods of patient-reported outcomes, particularly when care has been taken to minimize the differences between modes with respect to wording, interpretation, and response options [25,26]. Providing choice via a mixed mode study design has been shown to increase motivation among some participants [27] as well as lead to higher response rates [28] and a more representative study sample makeup [29].

Web-based survey acceptability was not affected by the presence of frailty or health comorbidities. Frailty was also not associated with preference or interest in technology use in a previous study of older adults [30]. Indeed, we observed that participants who reported one or more limitations in IADL were much more likely to prefer completing the questions in the future using a computer or tablet than to report *no preference*. A study that evaluated older adults' use of technology for personal or health-related tasks found a higher prevalence of use among those with certain health issues (pain and breathing problems) that limited activities of daily living. They concluded that technology might be useful to enhance communication and completion of tasks by removing some barriers associated with completing these tasks offline [7]. However, other research on older adults who report not using or discontinuing the use of web-based technologies cite functional impairments such as arthritis and visual deficits as reasons for nonuse [6,31]. Although future research is needed, it appears that web-based technology may facilitate research data collection for older adults with specific health-related limitations.

The findings of this pilot study indirectly inform a longer-term goal of incorporating remotely completed web-based assessments into A5322. In addition to age- or cognitive-based limitations with completing web-based assessments, remote completion also requires internet/smartphone access at home as well as the skills needed to independently access and submit web-based surveys. Given these caveats, the option of remote data completion may increase retention among participants who find it difficult to maintain regular clinic visits because of time, travel, or health restrictions. Although studies that rely solely on remote data collection are likely to have high attrition rates [32,33], a study design that incorporates remote web-based data collection while retaining the option of in-person study visits has been suggested as a method that can improve long-term retention [1].

Limitations

This pilot study was limited to participants whose primary language was English, and the design required that the web-based questionnaires be completed on site. Therefore, the results might not be generalizable to non–English-speaking individuals or to those without adequate computer or tablet or internet access. We were unable to have participants complete the web-based and paper-and-pencil questionnaires at the same time and therefore were not able to compare frequencies of behaviors by format because we could not assume that the frequency of these behaviors would be stable over time. We did not take within-person correlations into account when comparing the proportion of evaluable responses across survey administration types. Many of the questions were designed to be answered only by participants who endorsed a leading question that triggered another question (eg, the leading question "in the past 6 months, have you had sex with another person?" would trigger subsequent questions on the type of partners and condom use). The number of participants answering each question differed by mode of administration because they were asked at different points in time, with different frequencies of behaviors reported. Therefore, a matched approach was not possible for all comparisons. However, we were able to use McNemar test for a subset of questions (those asked of all participants) to account for within-person correlations. In these situations, the McNemar P values were consistent with the chi-square P values (Multimedia Appendix 4).

Conclusions

We found that in a group of older adults living with HIV being followed in a longitudinal observational study, completion of a web-based survey of questionnaires assessing sexual, substance use, and physical activity behaviors was perceived to be highly acceptable and feasible, with data quality on average being higher with the web-based versus paper-and-pencil format.

As persons living with HIV continue to age, often with comorbidities and disabilities, long-term participation in research will become more challenging, even as the need to understand their health-related outcomes continues to grow [34]. Web-based technologies can be a useful tool for valid data collection and may be a way to optimize the retention of these individuals.

Multimedia Appendix 1
Web-based survey questions with ≥1 rather not answer response.
[DOCX File , 13 KB - jmir_v22i11e18588_app1.docx]

Multimedia Appendix 2
Rather not answer responses by acceptability and feasibility.
[DOCX File , 13 KB - jmir_v22i11e18588_app2.docx]

Multimedia Appendix 3
Evaluable responses by questionnaire format (comparing web-based surveys with paper-and-pencil questionnaires from all A5322 participants).
[DOCX File , 15 KB - jmir_v22i11e18588_app3.docx]

Multimedia Appendix 4
Comparison of evaluable responses—chi-square test versus McNemar test.
[DOCX File , 14 KB - jmir_v22i11e18588_app4.docx]

References

1. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
2. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol 2006 Feb 1;163(3):197-203. [doi: 10.1093/aje/kwj036] [Medline: 16339049]
3. Bonk J. A road map for the recruitment and retention of older adult participants for longitudinal studies. J Am Geriatr Soc 2010 Oct;58 Suppl 2:S303-S307. [doi: 10.1111/j.1532-5415.2010.02937.x] [Medline: 21029058]
4. Hanscom B, Lurie JD, Homa K, Weinstein JN. Computerized questionnaires and the quality of survey data. Spine (Phila Pa 1976) 2002 Aug 15;27(16):1797-1801. [doi: 10.1097/00007632-200208150-00020] [Medline: 12195074]
5. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003 Jan;57(1):68-73. [doi: 10.1136/jech.57.1.68] [Medline: 12490652]
6. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2020-02-24]
7. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
8. Baer A, Saroiu S, Koutsky LA. Obtaining sensitive data through the Web: an example of design and methods. Epidemiol 2002 Nov;13(6):640-645. [doi: 10.1097/00001648-200211000-00007] [Medline: 12410004]
9. Smith B, Smith TC, Gray GC, Ryan MA, Millennium Cohort Study Team. When epidemiology meets the Internet: web-based surveys in the millennium cohort study. Am J Epidemiol 2007 Dec 1;166(11):1345-1354. [doi: 10.1093/aje/kwm212] [Medline: 17728269]
10. Nahm E, Bausell B, Resnick B, Covington B, Brennan PF, Mathews R, et al. Online research in older adults: lessons learned from conducting an online randomized controlled trial. Appl Nurs Res 2011 Nov;24(4):269-275. [doi: 10.1016/j.apnr.2009.09.004] [Medline: 20974077]

11. Remillard ML, Mazor KM, Cutrona SL, Gurwitz JH, Tjia J. Systematic review of the use of online questionnaires of older adults. J Am Geriatr Soc 2014 Apr;62(4):696-705 [FREE Full text] [doi: 10.1111/jgs.12747] [Medline: 24635138]

12. Tan JY, Nguyen TT, Tabrisky A, Siedle-Khan R, Napoles AM. Mobile technology for healthy aging among older HIV-positive black men who have sex with men: qualitative study. JMIR Aging 2018 Dec 10;1(2):e11723 [FREE Full text] [doi: 10.2196/11723] [Medline: 31518233]

13. Olalla J, de Lomas JM, Márquez E, González FJ, del Arco A, de La Torre J, et al. Experience of using an app in HIV patients older than 60 years: pilot program. JMIR Mhealth Uhealth 2019 Mar 6;7(3):e9904 [FREE Full text] [doi: 10.2196/mhealth.9904] [Medline: 30839281]

14. Harshbarger C, Burrus O, Zulkiewicz BA, Ortiz AM, Galindo CA, Garner BR, et al. Implementing web-based interventions in HIV primary care clinics: pilot implementation evaluation of positive health check. JMIR Form Res 2019 Apr 18;3(2):e10688 [FREE Full text] [doi: 10.2196/10688] [Medline: 30998219]

15. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and acceptability of ecological momentary assessment of daily functioning among older adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840 [FREE Full text] [doi: 10.1016/j.jagp.2016.11.019] [Medline: 28065496]

16. Alford K, Vera JH. Cognitive impairment in people living with HIV in the ART era: a review. Br Med Bull 2018 Sep 1;127(1):55-68. [doi: 10.1093/bmb/ldy019] [Medline: 29868901]

17. Erlandson KM, Schrack JA, Jankowski CM, Brown TT, Campbell TB. Functional impairment, disability, and frailty in adults aging with HIV-infection. Curr HIV/AIDS Rep 2014 Sep;11(3):279-290 [FREE Full text] [doi: 10.1007/s11904-014-0215-y] [Medline: 24966138]

18. Gallant J, Hsue PY, Shreay S, Meyer N. Comorbidities among US patients with prevalent HIV infection: a trend analysis. J Infect Dis 2017 Dec 19;216(12):1525-1533. [doi: 10.1093/infdis/jix518] [Medline: 29253205]

19. Erlandson KM, Perez J, Abdo M, Robertson K, Ellis RJ, Koletar SL, et al. Frailty, neurocognitive impairment, or both in predicting poor health outcomes among adults living with human immunodeficiency virus. Clin Infect Dis 2019 Jan 1;68(1):131-138 [FREE Full text] [doi: 10.1093/cid/ciy430] [Medline: 29788039]

20. Tassiopoulos K, Abdo M, Wu K, Koletar SL, Palella FJ, Kalayjian R, et al. Frailty is strongly associated with increased risk of recurrent falls among older HIV-infected adults. AIDS 2017 Oct 23;31(16):2287-2294 [FREE Full text] [doi: 10.1097/QAD.0000000000001613] [Medline: 28991026]

21. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

22. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]

23. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]

24. Mauz E, von der Lippe E, Allen J, Schilling R, Müters S, Hoebel J, et al. Mixing modes in a population-based interview survey: comparison of a sequential and a concurrent mixed-mode design for public health research. Arch Public Health 2018;76:8. [doi: 10.1186/s13690-017-0237-1] [Medline: 29423220]

25. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]

26. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 7;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]

27. Baines AD, Partin MR, Davern M, Rockwood TH. Mixed-mode administration reduced bias and enhanced poststratification adjustments in a health behavior survey. J Clin Epidemiol 2007 Dec;60(12):1246-1255. [doi: 10.1016/j.jclinepi.2007.02.011] [Medline: 17998079]

28. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. Eval Rev 2009 Oct;33(5):464-480. [doi: 10.1177/0193841X09340214] [Medline: 19605623]

29. Fitzgerald D, Hockey R, Jones M, Mishra G, Waller M, Dobson A. Use of online or paper surveys by Australian women: longitudinal study of users, devices, and cohort retention. J Med Internet Res 2019 Mar 14;21(3):e10672 [FREE Full text] [doi: 10.2196/10672] [Medline: 30869647]

30. Keränen NS, Kangas M, Immonen M, Similä H, Enwald H, Korpelainen R, et al. Use of information and communication technologies among older people with and without frailty: a population-based survey. J Med Internet Res 2017 Feb 14;19(2):e29 [FREE Full text] [doi: 10.2196/jmir.5507] [Medline: 28196791]

31. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 2;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]

XSL•FO

RenderX

32.   Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [Medline: 30952390]

33.   Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

34.   Poli A, Kelfve S, Motel-Klingebiel A. A research tool for measuring non-participation of older people in research on digital health. BMC Public Health 2019 Nov 8;19(1):1487 [FREE Full text] [doi: 10.1186/s12889-019-7830-x] [Medline: 31703655]

Abbreviations

ACTG: AIDS Clinical Trials Group
IADL: Instrumental Activities of Daily Living
NIH: National Institutes of Health
OR: odds ratio

XSL·FO
RenderX

<u>Original Paper</u>

Predictors of Internet Health Information–Seeking Behaviors Among Young Adults Living With HIV Across the United States: Longitudinal Observational Study

Warren Scott Comulada[1,2], DrPH; Mary Step[3], PhD; Jesse B Fletcher[4], PhD; Amanda E Tanner[5], MPH, PhD; Nadia L Dowshen[6], MD, MSHP; Sean Arayasirikul[7], PhD; Kristin Keglovitz Baker[8], PA-C, AAHIVS; James Zuniga[8], BA; Dallas Swendeman[1,2], PhD; Melissa Medich[9], PhD; Uyen H Kao[10], MPH; Adam Northrup[10], MS; Omar Nieto[10], BA; Ronald A Brooks[10], PhD; Special Projects Of National Significance Social Media Initiative Study Group[11]

[1]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States

[2]Department of Health Policy and Management, University of California, Los Angeles, Los Angeles, CA, United States

[3]College of Public Health, Kent State University, Kent, OH, United States

[4]Friends Research Institute, Inc, Los Angeles, CA, United States

[5]Department of Public Health, University of North Carolina Greensboro, Greensboro, NC, United States

[6]Division of Adolescent Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[7]Department of Pediatrics, University of California, San Francisco, San Francisco, CA, United States

[8]Howard Brown Health Center, Chicago, IL, United States

[9]Center for the Study of Healthcare Innovation, Implementation and Policy, Health Services Research & Development, Veterans Affairs Greater Los Angeles Health Care System, US Department of Veteran Affairs, Los Angeles, CA, United States

[10]Department of Family Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[11]See Acknowledgments

Corresponding Author:
Warren Scott Comulada, DrPH
Department of Psychiatry and Biobehavioral Sciences
University of California, Los Angeles
10920 Wilshire Blvd, Suite 350
Los Angeles, CA, 90024
United States
Phone: 1 310 794 0938
Email: wcomulada@mednet.ucla.edu

Abstract

Background: Consistent with young adults' penchant for digital communication, young adults living with HIV use digital communication media to seek out health information. Understanding the types of health information sought online and the characteristics of these information-seeking young adults is vital when designing digital health interventions for them.

Objective: This study aims to describe characteristics of young adults living with HIV who seek health information through the internet. Results will be relevant to digital health interventions and patient education.

Methods: Young adults with HIV (aged 18-34 years) self-reported internet use during an evaluation of digital HIV care interventions across 10 demonstration projects in the United States (N=716). Lasso (least absolute shrinkage and selection operator) models were used to select characteristics that predicted whether participants reported seeking general health and sexual and reproductive health (SRH) information on the internet during the past 6 months.

Results: Almost a third (211/716, 29.5%) and a fifth (155/716, 21.6%) of participants reported searching for general health and SRH information, respectively; 26.7% (36/135) of transgender young adults with HIV searched for gender-affirming care topics. Areas under the curve (>0.70) indicated success in building models to predict internet health information seeking. Consistent with prior studies, higher education and income predicted health information seeking. Higher self-reported antiretroviral therapy adherence, substance use, and not reporting transgender gender identity also predicted health information seeking. Reporting a sexual orientation other than gay, lesbian, bisexual, or straight predicted SRH information seeking.

XSL•FO
RenderX

Conclusions: Young adults living with HIV commonly seek both general health and SRH information online, particularly those exploring their sexual identity. Providers should discuss the most commonly sought SRH topics and the use of digital technology and be open to discussing information found online to better assist young adults with HIV in finding accurate information. Characteristics associated with health information–seeking behavior may also be used to develop and tailor digital health interventions for these young adults.

Introduction

Interventions for young adults living with HIV (YALH) increasingly capitalize on the popularity and integration of digital communication media into daily routines. Social media use has saturated the information landscape in the United States, with near-ubiquitous social media platform use among those younger than 30 years [1]. Consistent with their heavy use of digital forms of communication, youth and young adults use digital communication media to seek out personalized and pertinent health information [2]. Though an income-based digital divide persists [3], reliance on electronic health information is common among youth from marginalized populations, including those living with HIV [4] and unstably housed [5]. In fact, almost half (47%) of runaway and homeless youth sought information about HIV or other STIs, and 40% sought information about sex or sexuality from online sources [5].

While the penchant for digital communication among youth and young adults can be harnessed to develop digital health interventions (eg, interventions that provide health information through online sources, social media, and text messages), these strategies are challenged by the growing levels of health misinformation available from digital sources [6]. Younger people often need guidance engaging with accurate health information designed for them. Understanding the types of health information sought online and the characteristics of these information-seeking young adults, especially vulnerable or stigmatized populations, is vital when designing digital health interventions to reach these communities.

Prior studies on internet health information seeking have focused on the general population and patient populations not living with HIV, mostly adults. These studies demonstrated that certain characteristics, including greater socioeconomic stability [7,8], more internet experience [9], female gender [8], less perceived social support [10], better health care provider relationships, and greater health engagement are associated with greater online health information seeking [11]. Behavioral characteristics, including poorer mental or physical health and alcohol and tobacco use also correlate with internet health information seeking [7,8,12]. Importantly, Mitchell et al [13] found that sexual minority youth are more likely to use the internet to seek sexual health information than their heterosexual counterparts.

In one notable addition to the literature, Calvert et al [14] evaluated adults living with HIV and found that greater socioeconomic stability was associated with greater engagement with online health information seeking. Studies focused specifically on how YALH seek health information are needed because the stigma associated with HIV is a known barrier to care [15]. Stigma and the intersectionality of multiple marginalized identities (racial/ethnic minority, gender identity, sexual orientation) of many YALH may represent opportunities for safe exploration and information seeking through digital spaces [16]. Furthermore, a robust examination of predictive models of internet health-seeking behaviors will provide valuable information for tailoring digital interventions to YALH.

To address this goal, analyses for this paper applied machine learning (ML) methods to data from a digital health intervention initiative for YALH to identify salient characteristics that predict internet health-seeking behaviors. The comprehensive model of information seeking [17] and correlates of online health seeking from prior studies provided a framework for selecting candidate predictors. An individual's sociodemographic characteristics were conceptualized as preceding, and even influencing, where people seek information [18]. Information seeking refers to intentional efforts made by individuals to satisfy their information needs or goals [18], such as HIV-related health care needs. As a secondary aim, the study evaluated individual predictors selected by ML methods and compared these findings with extant literature.

Methods

Participants

Data used for this analysis were collected as part of an initiative funded by the Health Resources and Services Administration to evaluate digital health interventions targeting young people living with HIV (aged 13 to 34 years) across 10 demonstration sites in the United States. Digital interventions were developed by each site and varied in content and delivery format, which included automated text messaging, mobile apps, and social media. Common intervention elements included health promotion messages, enhanced communication with intervention staff using digital communication tools, and HIV medication and medical care appointment reminders. All interventions targeted HIV care continuum outcomes and were evaluated over 18 months.

Young people living with HIV were recruited in the following cities from October 2016 through May 2018: Chicago, Illinois; Cleveland, Ohio; Corpus Christi, Texas; Hershey, Pennsylvania; Los Angeles, California; New York, New York; Philadelphia, Pennsylvania; San Francisco, California; St Louis, Missouri;

XSL•FO

RenderX

and Winston-Salem, North Carolina. Recruitment took place through community and university clinics, health departments, a hospital system, and a community research site. Eligibility for study enrollment required young people to have a confirmed HIV diagnosis, be between the ages of 13 and 34 years, be capable of filling out audio computer-assisted self-interview (ACASI) assessments administered in English or Spanish, and meet at least one of the following criteria based on the US Department of Health and Human Services (HHS) common core indicators for monitoring HHS-funded HIV care services: (1) newly diagnosed with HIV within the last year upon enrollment, (2) not newly diagnosed and not currently engaged in HIV care, (3) never linked to HIV medical care, regardless of the duration of HIV infection, and (4) not virally suppressed, defined as having a viral load of 200 copies/mL or greater. Demonstration sites had additional eligibility criteria, such as being a patient at the site's clinic or owning a smartphone, if required by their digital health intervention. Participants from

all genders, races and ethnicities, and sexual orientations were included in the initiative. Details on the initiative and the intervention typology across sites are described in Medich et al [19].

Figure 1 shows the process used to select participants for analysis. Analyses in this paper incorporated predictors measured at baseline and outcomes measured at 6 months post enrollment (N=720 participants). Participants with missing baseline or 6-month assessments were excluded. Missing data occurred from errors saving electronic assessment files or missing assessments (eg, due to attrition after the baseline assessment). There were only 4 participants younger than 18 years old, making it difficult to model health-seeking behaviors in this group. Moreover, the younger participants represented a different patient population in terms of clinical practice. Therefore, they were excluded, and the final analytical sample contained 716 YALH.

Figure 1. Sample selection process for analyses using baseline data to predict internet health information–seeking behaviors assessed 6 months after baseline. YALH: young adults living with HIV.
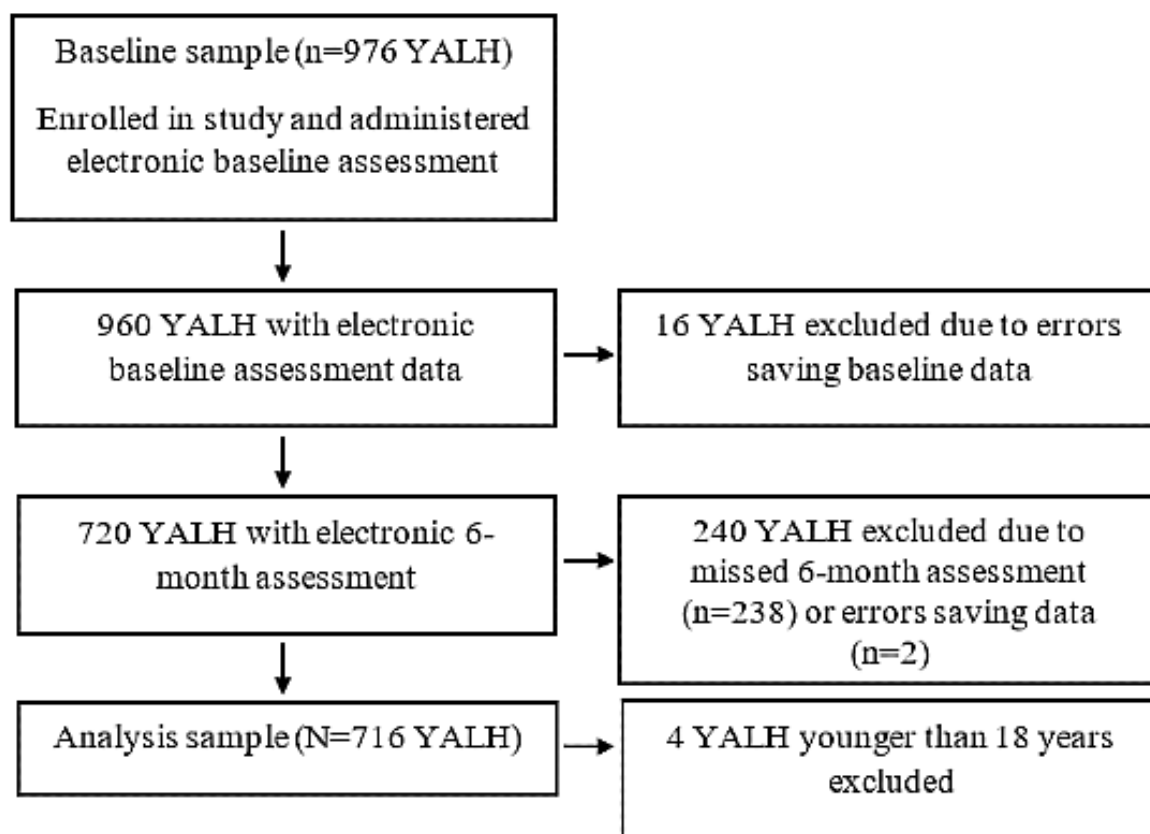
Procedure

All data collection procedures for the cross-site evaluation were approved by the institutional review board at the University of California, Los Angeles (UCLA; No. 15-001625), the institution that was responsible for collecting and evaluating data across the sites. At enrollment, each site screened, consented, and administered a baseline ACASI assessment to participants using Questionnaire Design Studio software (Nova Research Company). Sites also collected participants' medical chart data, either by hand abstraction or from administrative records associated with the receipt of Ryan White HIV/AIDS Program

funds. ACASI assessments were administered and medical chart data were obtained by sites every 6 months over the 18-month follow-up period. Sites submitted deidentified ACASI and medical chart data to the UCLA evaluation center through a web-based secure portal.

Measures

Measures treated as predictors were assessed at baseline. After baseline assessment began, measures that better captured evolving trends in technology usage among lesbian, gay, bisexual, transgender, and queer or questioning youth than baseline measures were developed and added to the 6-month

follow-up assessment. These measures are treated as outcomes in the analyses.

Predictors Measured at Baseline

Sociodemographic Characteristics

Age was calculated from the self-reported month and year of birth. Participants were asked to specify the race with which they identified and indicate whether they were Hispanic or Latinx. They designated their current gender identity with categories for male, female, transgender man, transgender woman, genderqueer or nonconforming, or other gender identity. Participants were also asked to categorize their sexual orientation as straight, lesbian or gay, bisexual, queer, other, or don't know/not sure; responses indicating "other" varied, included pansexual, nonsexual, and refusals to answer. Participants specified whether they were currently in school and the highest level of education they had completed. They were also asked to report monthly income "from all sources combined" and their current employment status (eg, full-time, part-time, student, or disabled). Housing stability was assessed by asking participants to indicate which type of place they stayed in the most in the past week (eg, a house or homeless shelter).

Region

Most transgender women were recruited in Los Angeles, since the Los Angeles site intervention targeted transgender women. Collinearity that would have resulted by including site and gender identity as predictors was addressed by replacing site with a predictor based on Census Bureau regions for the United States. Categories were created for the West (Los Angeles and San Francisco, California), Midwest (Chicago, Illinois; Cleveland, Ohio; and St Louis, Missouri), South (Corpus Christi, Texas, and Winston-Salem, North Carolina), and Northeast regions (Hershey and Philadelphia, Pennsylvania, and New York, New York).

Health Insurance

Participants were asked what type of health insurance they had. Insurance status was dichotomized as being insured versus not being insured or not knowing one's insurance status.

Time Since HIV Diagnosis

Time since HIV diagnosis was calculated as the number of years between the self-reported HIV diagnosis date and the baseline assessment date.

Antiretroviral Therapy Adherence

Antiretroviral therapy (ART) adherence was assessed using the self-rating scale item [20,21]. Categories were collapsed to indicate low ("very poor" or "poor"), neutral ("fair"), or high adherence ("good", "very good", or "excellent").

Viral Load

Viral load data were obtained via abstraction from patient medical records. Viral load was categorized as suppressed (at less than 200 copies/mL), unsuppressed, or missing. A missing data category was included because sites were unable to obtain medical chart data on all participants.

Doctor's Office Visitation

Similar to viral load, HIV-related ambulatory care visit attendance was obtained from medical record data. Attendance was categorized as having had an HIV medical visit in the past 6 months, not having had an HIV visit in the past 6 months, or missing.

Health Engagement

Health engagement was also assessed through the youth engagement with health services (YEHS) survey [22]. Responses were summed across 2 YEHS subscales for health access literacy (5 items; Cronbach α=.81) and health self-efficacy (5 items; Cronbach α=.89).

Provider Empathy

The consultation and relational empathy measure was used to assess participants' perceptions of health care provider empathy (10 items; Cronbach α=.98) [23].

Substance Use

Participants were asked to indicate any nonprescribed substances they used but did not inject in the past 6 months from a checklist (ie, recent use). Both proper names and street names of substances were presented in the checklist, such as methamphetamine and "Tina." Indicator variables were created to denote use (1) or nonuse (0) for alcohol, tobacco, marijuana, and other substances, such as synthetic marijuana, methamphetamine, cocaine, heroin, and painkillers. Other substances were not modeled separately due to self-reported rates of use that were less than 10%, except for methamphetamines (118/720, 16.3%), inhalants (113/720, 15.7%), and powder cocaine (88/720, 12.2%). Participants were asked about lifetime and recent injection drug use, excluding prescribed medications.

Perceived Confidence in Receiving Social Support From Family and Friends

Perceived social support availability from family and friends was assessed through 3 social support items from the coping self-efficacy scale (3 items; Cronbach α=.83) [24], in which participants were asked about confidence in receiving support from family and friends on a scale from 0 (not confident at all) to 10 (very confident).

HIV Status Disclosure

HIV status disclosure was dichotomized as disclosure to one or more individuals or to none based on the participant response to having ever told anyone that they have HIV. If they had disclosed their status, they were asked to indicate types of individuals to whom they disclosed their HIV status (eg, partners and family members).

HIV-Related Stigma

HIV-related stigma was assessed through the revised HIV stigma scale (10 items; Cronbach α=.89) [25]. Using a scale from 1 (strongly disagree) to 4 (strongly agree), respondents were asked to rate their agreement with statements about experiencing HIV stigma.

General Physical and Mental Health

General physical and mental health quality of life was assessed with 4 questions from the 12-item Short-Form Health Survey [26]. Participants were asked if they "felt calm and peaceful," had "a lot of energy," or "felt downhearted and blue" over the past 4 weeks. The 3 items were summed to create a mental health measure (Cronbach α=.66). Participants were also asked how often their physical health or emotional problems interfered with social activities.

The Media and Technology Usage and Attitudes Scale

The Media and Technology Usage and Attitudes Scale (MTUAS) subscales [27] were administered to assess the frequency of digital communication use, including emailing (4 items; Cronbach α=.88), texting (4 items; Cronbach α=.67), smartphone use (9 items; Cronbach α=.90), internet searching (4 items; Cronbach α=.91), and general social media use (9 items; Cronbach α=.92). Subscales on positive attitudes toward technology (6 items; Cronbach α=.88), anxiety about being without technology or dependence on technology (3 items; Cronbach α=.88), and negative attitudes toward technology (3 items; Cronbach α=.81) were also administered.

Outcomes Measured at 6 Months Post Enrollment

Participants were asked what types of digital media and communication tools they used and what types of information were sought and discussed. For this analysis, the focus centered on questions that queried the types of information that were sought through the internet. Sexual health information (eg, practicing safer sex and HIV information) discussed or sought through text messaging, email, private messaging, and social networking applications is also presented to describe the sample. Two binary outcome measures were created for (1) having looked up general health (GH) information on the internet in the past 6 months and (2) having looked up sexual and reproductive health (SRH) information on the internet in the past 6 months. Transgender health information seeking (eg, gender-affirming hormone information) was also assessed, but rates were too low to analyze using ML models.

Statistical Analysis

All analyses were conducted using R software (version 3.5.3; R Project for Statistical Computing) [28]. Data were randomly split into training (537/720, 74.6%) and testing data sets (179/720, 24.9%). A ML approach was chosen to meet the aims of the paper to build a predictive model and evaluate individual predictors selected by the model. In this vein, we used lasso (least absolute shrinkage and selection operator) regression as

the ML approach because it fits a model to all candidate predictors and shrinks regression coefficients to zero for predictors that do not adequately contribute to error minimization. In other words, lasso regression provides a distinguishable subset of predictors, in contrast to ridge regression, which does not constrain regression coefficients to be zero, or to other ML approaches that provide less interpretable parameter estimates, such as random forest algorithms. The glmnet R package [29] was used to fit lasso logistic regression models to the training data set using 10-fold validation to select predictors for seeking general health information and SRH information via the internet.

Accuracy of the lasso models was gauged by using parameter estimates to predict internet health information seeking in the test data and comparing predictions to observed outcome values. Receiver operating curves (ROCs) were plotted to evaluate the sensitivity and specificity of predictions over a range of probability thresholds. Areas under the ROC curve (AUCs) are presented to gauge the accuracy of predictions. An AUC of 0.50 indicates a model that performs no better than chance.

Traditional logistic regressions were fit to the training data using predictors selected by lasso models to aid interpretation of associations between predictors and internet health seeking. Odds ratios (ORs) are reported. Statistical significance levels are not presented due to difficulties interpreting regression coefficient P values for subsets of predictors selected using ML algorithms.

Results

Sample Characteristics

Tables 1 and 2 show variables that were evaluated as candidate predictors of seeking health information on the internet across the 10 demonstration sites (N=716). Two-thirds of the participants were aged 25 to 34 years (483/716, 67.5%). Half of the participants reported a non-Latinx African American racial/ethnic identity (362/716, 50.6%); 27.9% (200/716) reported Latinx ethnicity. Most participants reported male gender (506/716, 70.7%). Nearly one-fifth (130/716, 18.2%) of participants identified as transgender women, and 5 of the 716 participants identified as transgender men (.01%). Approximately half of the participants identified as gay or lesbian (393/716, 54.9%), and half reported having no more than a high school education and access to stable housing (368/716, 51.4% and 363/716, 50.7%, respectively). The median monthly income was US $800 (IQR US $200 to $1500).

Table 1. Sociodemographic and HIV-related health care measures entered into lasso models as internet health-seeking predictors.

Characteristic	Values
Site, n (%)	
Chicago	84 (11.7)
Cleveland	82 (11.5)
Corpus Christi	86 (12.0)
Hershey	36 (5.0)
Los Angeles	110 (15.4)
New York	30 (4.2)
Philadelphia	28 (3.9)
St Louis	84 (11.7)
San Francisco	88 (12.3)
Winston-Salem	88 (12.3)
Age group (years), n (%)	
18-24	233 (32.5)
25-34	483 (67.5)
Race/ethnicity, n (%)	
Latinx	200 (27.9)
Non-Latinx African American	362 (50.6)
Non-Latinx White	121 (16.9)
Other racial/ethnic identity	33 (4.6)
Gender identity, n (%)	
Male (ie, cisgender man)	506 (70.7)
Female (ie, cisgender woman)	57 (8.0)
Transgender-identified	135 (18.9)
Other gender identity	18 (2.5)
Sexual orientation, n (%)	
Straight	168 (23.5)
Gay or lesbian	393 (54.9)
Bisexual	99 (13.8)
Other sexual orientation	56 (7.8)
Education, n (%)	
High school/GED[a] or less	368 (51.4)
Some college education	220 (30.7)
College degree or trade certification	128 (17.9)
Current residence, n (%)	
Stable housing	363 (50.7)
Unstable housing[b]	353 (49.3)
Monthly income, median (IQR)[c]	800 (200-1500)
Health insurance status, n (%)	
Insured	489 (68.3)
Not insured or don't know	227 (31.7)
HIV diagnosis, n (%)	

Characteristic	Values
Within past 12 months	229 (32.0)
>12 months	483 (67.5)
Don't know	4 (0.6)
ART[d] adherence, n (%)	
High	362 (50.6)
Neutral	67 (9.4)
Low	57 (40.1)
Not on ART	230 (32.1)
Viral suppression[e], n (%)	270 (42.3)
Recent doctor's office visit[f], n (%)	557 (86.9)
Youth health engagement, mean (SD)	
Health access literacy (1-16[g])	12.0 (3.5)
Health self-efficacy (2-20[g])	16.8 (4.0)
Provider empathy (10-50[g]), mean (SD)	42.6 (9.8)

[a]GED: general education development.

[b]Unstably housed group includes 6 participants who reported being hospitalized or in prison.

[c]N=655 due to "don't know/not sure" responses.

[d]ART: antiretroviral therapy.

[e]Viral suppression information obtained from medical chart data (N=638).

[f]HIV doctor's visit information obtained from medical chart data (N=641).

[g]Minimum and maximum values for scales are shown in parentheses.

Two-thirds of the participants reported being diagnosed with HIV for over a year (483/716, 67.5%) and having health insurance (489/716, 68.3%). Based on medical chart data, most participants had visited a doctor's office for HIV in the past 6 months (557/641, 86.9%), but less than half of the participants were virally suppressed (270/641, 42.1%). Most participants disclosed their HIV status to someone (654/716, 91.3%), including partners (444/716, 62.0%) and friends or family (551/716, 77.0%). Two-thirds of the participants drank alcohol (490/716, 68.4%) and a little over half used marijuana (422/716, 58.9%) and tobacco (373/716, 52.1%) within the past 6 months. Less than half of the participants reported using other noninjected substances (320/716, 44.7%). Most participants used alcohol or at least one noninjectable substance within the past 6 months (598/716, 83.5%). A total of 16.2% (115/711) of the participants reported injection drug use.

Almost a third of the participants reported searching for GH topics on the internet over the past 6 months (211/716, 29.5%), followed by SRH information (155/716, 21.6%). In regard to discussing or seeking SRH through other digital communication tools, rates were 20.0% (143/716) for text messaging, 10.2% (73/716) for email, 8.7% (62/716) for social networking apps (eg, Facebook or Instagram), and 7.8% (56/716) for private messaging. Nearly a quarter of transgender participants reported searching for transgender-specific topics (36/135, 26.7%), including hormone therapy (30/135, 22.2%) and surgeries (24/135, 17.8%).

XSL•FO

RenderX

Table 2. Substance use, HIV-related disclosure and stigma, physical and mental health, and Media and Technology Usage and Attitudes Scale measures entered into lasso models as internet health-seeking predictors.

Characteristic	Values
Recent substance use, n (%)	
Alcohol	490 (68.4)
Tobacco	373 (52.1)
Marijuana	422 (58.9)
Other noninjected drugs	320 (44.7)
Lifetime injection drug use[a], n (%)	115 (16.2)
HIV status disclosure, n (%)	654 (91.3)
HIV-related stigma (10-40[b]), mean (SD)	24.5 (7.5)
Perceived social support (0-30[b]), mean (SD)	21.0 (7.7)
Physical and mental health, mean (SD)	
Mental health (3-18[b])	9.4 (3.6)
Mental and physical health (1-6[b])	4.0 (1.7)
MTUAS[c], mean (SD)	
Email usage (0-36[b])	19.8 (11.0)
Text messaging (0-27[b])	21.4 (6.0)
Smartphone usage (0-81[b])	62.1 (18.3)
Internet searching (0-81[b])	27.2 (9.8)
General social media usage (0-81[b])	49.1 (25.0)
Positive technology attitudes (6-30[b])	24.0 (5.1)
Anxiety without or dependent on technology (3-15[b])	10.9 (3.4)
Negative technology attitudes (3-15[b])	9.4 (3.1)

[a]N=711 for injected drugs due to refusal responses.

[b]Minimum and maximum values for scales shown in parentheses.

[c]MTUAS: Media and Technology Usage and Attitudes Scale.

Predictors of Internet Health Information Seeking

Figure 2 shows the ROCs for lasso-based predictions of GH and SRH information seeking in the test data. Curves above the 45° line indicate a degree of predictability beyond chance. AUCs for lasso models fit to GH information–seeking and SRH information–seeking outcomes are 0.76 and 0.73, respectively. To aid interpretation, we describe the accuracy of the model for a probability threshold of 0.50, where we classified participants as having searched the internet for health information if the predicted probability was greater than 0.50. A total of 32.4% (58/179) of the participants in the test data sought GH information on the internet. We correctly classified 16 as having sought GH information and correctly classified 112 of the 121 participants who did not seek GH information. Based on a 0.50 threshold, the accuracy of the GH model was (16 + 112) / 179

= 71.5%. Using the same formula, the accuracy of the SRH model at the 0.50 threshold was 70%.

Table 3 and 4 show ORs from logistic models fit to internet GH information–seeking and SRH information–seeking outcomes. Covariates are predictors selected from each lasso model. Mostly consistent with our hypotheses, having a high school degree or less was associated with lower odds of seeking GH and SRH information on the internet relative to having a higher degree (OR 0.49 and 0.68, respectively). Reporting high monthly income was associated with higher odds of seeking SRH information on the internet relative to no, low, or unreported monthly income. In a contradictory fashion, reporting low monthly income was associated with lower odds of seeking SRH information on the internet relative to no or unreported monthly income (OR 0.59). Participants reporting recent use of alcohol, tobacco, and marijuana had higher odds of seeking GH and SRH information (OR 1.29-1.70).

Figure 2. Receiver operating curves showing performance of lasso-selected predictors to predict general health information (left) and sexual and reproductive health information seeking (right) for different probability thresholds. Lasso: least absolute shrinkage and selection operator.
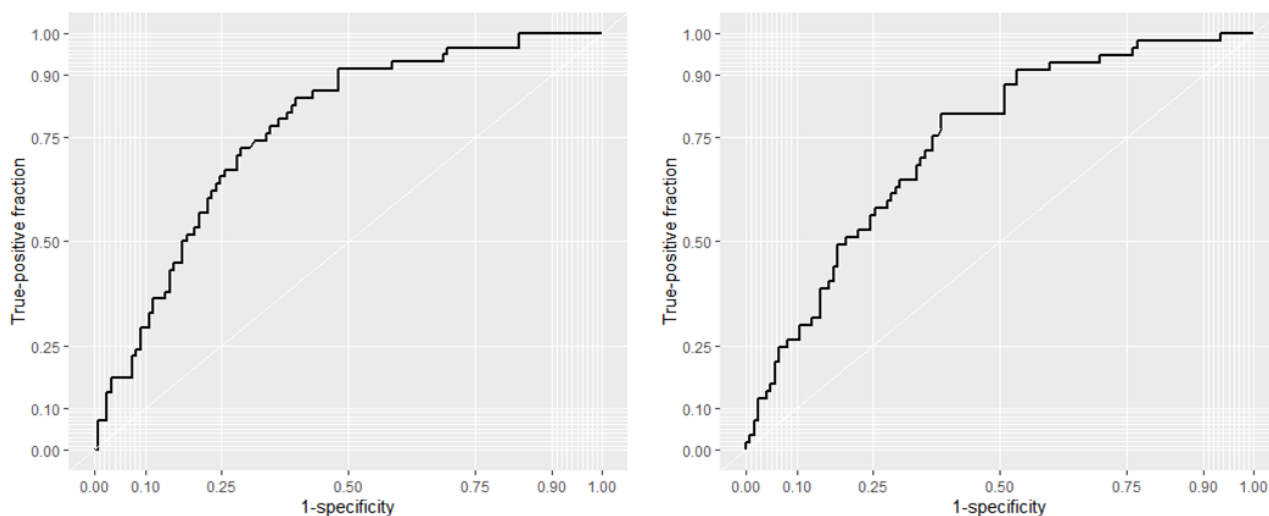
Table 3. Odds ratios from logistic regressions of general health information seeking through the internet over the past 6 months. Predictors were selected by lasso regression models.

Predictor	OR[a]
Gender	
Male	1.19
Transgender-identified	0.41
Female (cisgender woman) or other[b]	N/A[c]
Education	
High school or less	0.49
Some college, college degree, or trade certificate[b]	N/A
Monthly income	
Low (under median)	0.66
None, don't know, or high (median or above)	N/A
ART[d] adherence	
High	1.62
Not using, low, or neutral	N/A
Substance use[e]	
Alcohol	1.47
Tobacco	1.36
Marijuana	1.45

[a]OR: odds ratio.

[b]Reference category.

[c]N/A: not applicable.

[d]ART: antiretroviral therapy.

[e]Reference category was none for each substance.

Table 4. Odds ratios from logistic regressions of sexual and reproductive health information seeking through the internet over the past 6 months. Predictors were selected by lasso regression models.

Predictor	OR[a]
Gender	
Male	1.26
Transgender-identified	0.50
Female (cisgender woman) or other[b]	N/A[c]
Sexual orientation	
Other	2.48
Straight, gay, lesbian, or bisexual[b]	N/A
Race/ethnicity	
White	1.68
Latinx, non-Latinx, Black, or Other[b]	N/A
Education	
High school or less	0.68
College degree or trade certificate	1.45
Some college[b]	N/A
Monthly income	
Low (under median)	0.59
High (median or above)	1.18
None or don't know[b]	N/A
ART[d] adherence	
High	1.51
Not using, low, or neutral[b]	N/A
Substance use[e]	
Alcohol	1.29
Tobacco	1.62
Marijuana	1.70

[a]OR: odds ratio.

[b]Reference category.

[c]N/A: not applicable.

[d]ART: antiretroviral therapy.

[e]Reference category was none for each substance.

Self-reported high ART adherence was associated with higher odds of seeking GH and SRH information versus low adherence or not being on ART (OR 1.62 and 1.51, respectively). White ethnicity was associated with higher odds of seeking SRH information versus other racial/ethnic groups (OR 1.68). Male gender identity was associated with higher odds of seeking GH and SRH information (OR 1.19 and 1.26, respectively) and transgender gender identity was associated with lower odds of seeking GH and SRH information (OR 0.41 and 0.50, respectively) relative to other gender identities. The odds of seeking SRH information online were approximately twice as high for those reporting "other" as their sexual orientation (ie, excluding those identifying as gay, lesbian, bisexual, or straight) (OR 2.48).

Discussion

Principal Findings

This study is among the first to report internet health information–seeking behaviors among YALH. We found that a significant minority of YALH used the internet to find GH (211/716, 29.5%) and SRH information (155/716, 21.6%). The rates of technology use and health information seeking were similar in this population to previous reports of predominantly

racial/ethnic minority samples of homeless youth, who may face many similar challenges [30].

Patterns of seeking health information were associated with several demographic factors. As reported in the general population [8], YALH in this sample with higher socioeconomic status (ie, education, income) were more likely to go online to seek information regarding both GH and SRH. Interestingly, reporting a sexual orientation of "other" as opposed to gay, straight, or bisexual was also associated with increased SRH information seeking. This may reflect that adolescents and young adults who are exploring their sexuality may feel more comfortable finding health information online [31] as opposed to seeking health information from a person (eg, provider) due to perceived or enacted stigma in the health care setting [32].

Also consistent with general population findings [2], health-related information seeking in the sample was most likely to be directed toward GH topics, like diet and exercise. This focus likely stems from progress in HIV treatment and care [33] and highlights the importance of providers focusing on holistic health. Fewer YALH searched for SRH, most commonly to explore STI symptoms, testing, and treatment. Among transgender individuals (mostly transgender women), nearly a quarter searched for information about hormones, surgery, or other procedures. This is particularly important given poor access to gender-affirming services experienced by this population [34] and underscores the need for integration of gender-affirming care with HIV prevention and treatment services.

Limitations

Several study limitations should be noted. We attempted to engage young HIV-positive individuals who were struggling with adherence and engagement in care. However, this sample did not include those who are disengaged or lost to care, possibly due to syndemic health issues. This group may have very different internet health information–seeking patterns. It is also important to acknowledge that our sample was recruited to participate in digital HIV interventions, suggesting a higher proportion of YALH who seek health information on the internet than the general population of YALH. Further, while youth were recruited from 10 sites across the United States, regional differences in service options and HIV-related stigma may differentially affect YALH (ie, youth living in more rural areas). Region was not retained as a predictor in the final model, but

regional differences may not have been adequately captured by study site locations or the regional predictor variable that we created.

Conclusions

Despite these limitations, this is one of the first studies to address internet health information–seeking behaviors among a marginalized group of youth living with a chronic disease. High rates of internet use among YALH and nearly one-quarter of participants seeking health information online have important implications for clinicians and health educators working with YALH and other marginalized populations. Health care providers should receive training in how to engage in open discussions with patients about their technology use, the SRH topics they search for, and ways of ensuring the information being accessed online is reputable. These direct discussions may help reduce stigma and be particularly useful in supporting transitional age youth. Transitioning from pediatric to adult HIV care is commonly associated with poor retention in care [35]; leveraging eHealth literacy support represents an opportunity to improve care outcomes during this period.

While measures related to eHealth literacy (ie, YEHS and MTUAS subscales) were not retained in models, interventions to build transactional eHealth literacy skills (ie, skills to locate and understand, exchange, evaluate, and apply health information [36]) among YALH may still strengthen their engagement in care and increase access to high-quality health information via trusted communication channels (eg, governmental organizations), even through social networking platforms [37,38]. Though user-generated health information content shared on social networking platforms may not be as accurate or trustworthy as scientific or governmental sources, there is value in social networking platforms and tools regarding reach and engagement. Widening disparities in the quality of health information online, particularly on popular social media, may compromise the adoption of these platforms by trusted creators of online health information. For example, though the CDC maintains accounts on legacy social media networks (eg, Instagram), newer technologies can quickly lure younger adults away from carefully crafted messaging. To meet this growing need for trustworthy health information online that serves YALH, creators of digital health information need to be innovative in developing strategies for meeting YALH where they are online.

XSL·FO

RenderX

Conflicts of Interest

None declared.

References

1. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-05-07] [WebCite Cache ID 78C8HLwUK]

2. Rideout V, Fox S, Trust W. Digital health practices, social media use, and mental well-being among teens and young adults in the US. Articles, Abstracts, and Reports 2018;1093:1-96.

3. Anderson M, Kumar M. Digital divide persists even as lower-income Americans make gains in tech adoption. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2019-06-04] [WebCite Cache ID 78t0IXLyH]

4. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The Use of Mobile Health Applications Among Youth and Young Adults Living with HIV: Focus Group Findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260 [FREE Full text] [doi: 10.1089/apc.2016.0044] [Medline: 27214751]

5. Barman-Adhikari A, Rice E. Sexual Health Information Seeking Online Among Runaway and Homeless Youth. J Soc Social Work Res 2011 Jun;2(2):88-103 [FREE Full text] [doi: 10.5243/jsswr.2011.5] [Medline: 22247795]

6. Chou WYS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]

7. Gonzalez M, Sanders-Jackson A, Emory J. Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012. J Med Internet Res 2016 Jul 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]

8. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: Changes between 2002 and 2012. Health Informatics J 2016 Dec;22(4):804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]

9. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]

10. Kim SC, Shah DV, Namkoong K, McTavish FM, Gustafson DH. Predictors of Online Health Information Seeking Among Women with Breast Cancer: The Role of Social Support Perception and Emotional Well-Being. J Comput Mediat Commun 2013 Jan;18(2):98-118 [FREE Full text] [doi: 10.1111/jcc4.12002] [Medline: 24634575]

11. Kim SC, Shah DV, Namkoong K, McTavish FM, Gustafson DH. Predictors of Online Health Information Seeking Among Women with Breast Cancer: The Role of Social Support Perception and Emotional Well-Being. J Comput Mediat Commun 2013 Jan;18(2):98-118 [FREE Full text] [doi: 10.1111/jcc4.12002] [Medline: 24634575]

12. Nikoloudakis IA, Vandelanotte C, Rebar AL, Schoeppe S, Alley S, Duncan MJ, et al. Examining the Correlates of Online Health Information-Seeking Behavior Among Men Compared With Women. Am J Mens Health 2018 Sep;12(5):1358-1367 [FREE Full text] [doi: 10.1177/1557988316650625] [Medline: 27193765]

13. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157. [doi: 10.1093/her/cyt071] [Medline: 23861481]

14. Calvert JK, Aidala AA, West JH. An Ecological View of Internet Health Information Seeking Behavior Predictors: Findings from the CHAIN Study. Open AIDS J 2013;7:42-46 [FREE Full text] [doi: 10.2174/1874613601307010042] [Medline: 24222812]

15. Rao D, Frey S, Ramaiya M. eHealth for Stigma Reduction Efforts Designed to Improve Engagement in Care for People Living with HIV. Curr HIV/AIDS Rep 2018 Dec;15(6):397-402. [doi: 10.1007/s11904-018-0414-z] [Medline: 30232579]

16. DeHaan S, Kuper LE, Magee JC, Bigelow L, Mustanski BS. The interplay between online and offline explorations of identity, relationships, and sex: a mixed-methods study with LGBT youth. J Sex Res 2013;50(5):421-434. [doi: 10.1080/00224499.2012.661489] [Medline: 22489658]

17. Case DO, Given LM. Looking for information: a survey of research on information seeking, needs, and behavior. Bingley, UK: Emerald Group Publishing; 2016.

18. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. Health Commun 2007;22(2):153-167. [doi: 10.1080/10410230701454189] [Medline: 17668995]

19. Medich M, Swendeman DT, Comulada WS, Kao UH, Myers JJ, Brooks RA, Special Projects Of National Significance Social Media Initiative Study Group. Promising Approaches for Engaging Youth and Young Adults Living with HIV in HIV Primary Care Using Social Media and Mobile Technology Interventions: Protocol for the SPNS Social Media Initiative. JMIR Res Protoc 2019 Jan 31;8(1):e10681 [FREE Full text] [doi: 10.2196/10681] [Medline: 30702434]

XSL•FO

RenderX

20.   Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the single-item self-rating adherence scale for use in routine clinical care of people living with HIV. AIDS Behav 2013 Jan;17(1):307-318 [FREE Full text] [doi: 10.1007/s10461-012-0326-7] [Medline: 23108721]

21.   Lu M, Safren SA, Skolnik PR, Rogers WH, Coady W, Hardy H, et al. Optimal recall period and response task for self-reported HIV medication adherence. AIDS Behav 2008 Jan;12(1):86-94. [doi: 10.1007/s10461-007-9261-4] [Medline: 17577653]

22.   Sebastian RA, Ramos MM, Stumbo S, McGrath J, Fairbrother G. Measuring youth health engagement: development of the youth engagement with health services survey. J Adolesc Health 2014 Sep;55(3):334-340. [doi: 10.1016/j.jadohealth.2014.02.008] [Medline: 24709299]

23.   Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract 2004 Dec;21(6):699-705. [doi: 10.1093/fampra/cmh621] [Medline: 15528286]

24.   Chesney M, Neilands T, Chambers D, Taylor JM, Folkman S. A validity and reliability study of the coping self-efficacy scale. Br J Health Psychol 2006 Sep;11(Pt 3):421-437 [FREE Full text] [doi: 10.1348/135910705X53155] [Medline: 16870053]

25.   Wright K, Naar-King S, Lam P, Templin T, Frey M. Stigma scale revised: reliability and validity of a brief measure of stigma for HIV+ youth. J Adolesc Health 2007 Jan;40(1):96-98 [FREE Full text] [doi: 10.1016/j.jadohealth.2006.08.001] [Medline: 17185215]

26.   Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]

27.   Rosen LD, Whaling K, Carrier LM, Cheever NA, Rokkum J. The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Human Behav 2013 Nov 01;29(6):2501-2511 [FREE Full text] [doi: 10.1016/j.chb.2013.06.006] [Medline: 25722534]

28.   A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2020-10-16]

29.   Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]

30.   VonHoltz LAH, Frasso R, Golinkoff JM, Lozano AJ, Hanlon A, Dowshen N. Internet and Social Media Access Among Youth Experiencing Homelessness: Mixed-Methods Study. J Med Internet Res 2018 May 22;20(5):e184 [FREE Full text] [doi: 10.2196/jmir.9306] [Medline: 29789281]

31.   Steinke J, Root-Bowman M, Estabrook S, Levine DS, Kantor LM. Meeting the Needs of Sexual and Gender Minority Youth: Formative Research on Potential Digital Health Interventions. J Adolesc Health 2017 May;60(5):541-548 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.11.023] [Medline: 28108088]

32.   Hadland SE, Yehia BR, Makadon HJ. Caring for Lesbian, Gay, Bisexual, Transgender, and Questioning Youth in Inclusive and Affirmative Environments. Pediatr Clin North Am 2016 Dec;63(6):955-969 [FREE Full text] [doi: 10.1016/j.pcl.2016.07.001] [Medline: 27865338]

33.   Katz IT, Maughan-Brown B. Improved life expectancy of people living with HIV: who is left behind? Lancet HIV 2017 Aug;4(8):e324-e326 [FREE Full text] [doi: 10.1016/S2352-3018(17)30086-3] [Medline: 28501496]

34.   Padula WV, Baker K. Coverage for Gender-Affirming Care: Making Health Insurance Work for Transgender Americans. LGBT Health 2017 Aug;4(4):244-247. [doi: 10.1089/lgbt.2016.0099] [Medline: 28708447]

35.   Tanner AE, Philbin MM, Chambers BD, Ma A, Hussen S, Ware S, et al. Healthcare Transition for Youth Living With HIV: Outcomes from a Prospective Multi-site Study. J Adolesc Health 2018 Aug;63(2):157-165 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.02.004] [Medline: 29887488]

36.   Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a Transactional Model of eHealth Literacy: Concept Analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]

37.   Han H, Hong H, Starbird LE, Ge S, Ford AD, Renda S, et al. eHealth Literacy in People Living with HIV: Systematic Review. JMIR Public Health Surveill 2018 Sep 10;4(3):e64 [FREE Full text] [doi: 10.2196/publichealth.9687] [Medline: 30201600]

38.   Paige SR, Krieger JL, Stellefson ML. The Influence of eHealth Literacy on Perceived Trust in Online Health Communication Channels and Sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]

Abbreviations

ACASI: audio computer-assisted self-interview
ART: antiretroviral therapy
AUC: areas under the receiver operating curve
GH: general health
HHS: US Department of health and Human Services.
ML: machine learning

MTUAS: Media and Technology Usage and Attitude Scale
OR: odds ratio
ROC: receiver operating curve
SRH: sexual and reproductive health
UCLA: University of California, Los Angeles
YALH: young adults living with HIV
YEHS: youth engagement with health services

XSL•FO
RenderX

<u>Original Paper</u>

Expressions of Individualization on the Internet and Social Media: Multigenerational Focus Group Study

Gwendolyn Mayer[1], DipPsych; Simone Alvarez[2], Dr sc hum; Nadine Gronewold[1], MSc; Jobst-Hendrik Schultz[1], PD, Dr med

[1]Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg, Germany
[2]Texas A&M University - Central Texas, Killeen Texas, TX, United States

Corresponding Author:
Gwendolyn Mayer, DipPsych
Department of General Internal Medicine and Psychosomatics
Heidelberg University Hospital
Im Neuenheimer Feld 130 3
Heidelberg, 69120
Germany
Phone: 49 62215635685
Email: gwendolyn.mayer@med.uni-heidelberg.de

Abstract

Background: Growing individualization within the past decades has been described as a fundamental shift in society. Studies have reported how the digital age promotes new forms of individualism with self-tracking technologies and self-presentation in social networks. Potential harmful effects on the mental health of young adults have already been at the forefront of research. However, 2 questions that remain unanswered are how emotional experiences and expressions of self-relatedness differ among generations in their usage of the internet and social media, and if an increasing individualism can be observed by this.

Objective: The aim of this study is to examine whether the use of the internet and social media has led people to be more concerned about themselves than former generations. The potential consequences of mental and emotional distress among different age groups are analyzed.

Methods: A focus-group approach was chosen to study the following age groups: Baby Boomers (those born in 1950-1965), Generation X (those born in 1966-1980), and Digital Natives (those born in 1981-2000). We organized 6 focus groups with 36 participants who discussed their private usage of the internet and social media, different devices, platforms and functions, communication behavior, and self-tracking. We applied inductive category formation and followed the Standards for Reporting Qualitative Research (SRQR) checklist.

Results: We found differences in the 3 studied generations regarding the reasons for their use of the internet and social media, the effects of this use, personal feelings and experiences, expressions of self-relatedness, and social relationships. The Baby Boomers voiced a wish to stay autonomous while being in contact with their personal network. Generation X included enthusiastic members who appreciate self-tracking for curiosity and fascination, as well as people who felt fears about data surveillance. The Digital Natives reported a wish to optimize their own body by self-tracking while being faced with norms and expectations that were communicated via the internet and social media.

Conclusions: All generations expressed self-relatedness, yet by different means. The Baby Boomers expressed less individualism than Generation X and the Digital Natives, who felt the highest strain due to social comparisons. However, all generations reported specific, potentially problematic consequences for their mental health. Age-specific coping strategies are necessary to promote a mentally healthy way of using the internet and social media.

XSL•FO
RenderX

Introduction

Background

Individualization is considered to be a major transformation of postmodern society, and many efforts have been made to describe its reasons, preconditions, and consequences [1-3]. The beginning of its development started much earlier (likely with the courtly society of the Age of the Enlightenment [4]); however, most social science research focuses on the rise of individualization after World War II, characterized by an unprecedented number of choices opening up to individuals due to growing economic prosperity. Yet this freedom came along with uncertainty due to a loss of tradition, fragile norms, family structures, and gender roles [5]. The individual faced new challenges, such as having to plan one's own personal life in the realms of education, marriage, or residence, which was formerly decided upon by one's family or community [6]. This burden of choices has been linked to mental distress, with depression as its severest form [7].

Adolescents facing a modern paradigm of being as individualistic as possible have to cope with the increasing pressure of growing responsibilities. Being less tied to traditionally demarcated ways, they are now forced to make their own decisions at earlier ages than were former generations. Identity formation has become more than just a developmental step—it is a valuable resource needed for coping with the demands of individualization in a successful life [8,9].

Despite many theoretical approaches, a joint definition of the terms *individualization* and *identity,* as well as their interrelationships, is often missing. In this study, individualization is defined as an orientation of action that moves away from social or collective rules and toward person-based choices [10]. The precondition of this development is a stable and continuous feeling of being an individual, respectively, a *self* with an *identity* [11]. Role-theory describes the twofold nature of identity, with a social identity on the one hand, and a self-referential, personal identity on the other hand [12]. While personal identity results from one's own experienced biography and gives one a feeling of uniqueness, social identity is built on affiliations and relationships with the social environment [13]. In turn, a constant sense of not fulfilling the expectations of others leads to an effort, which has been called *self-optimization* [14,15].

Identity in the Digital Age

In the past 2 decades, mobile devices have allowed people to stay connected with others more than ever before. People are now interrelated in groups (eg, via apps such as WhatsApp) and permanently available at any place and time. This development impacts our identity in 2 ways: First, there is a technological acceleration of formerly long-lasting processes, whereby time pressure prevents us from keeping permanent, reliable relationships. As relationships are a constituting precondition of building up a stable, healthy identity, formerly lasting identities are transformed into "open, experimental, and fragmentary self-designs" [16]. Secondly, staying connected via electronic devices becomes part of our self-definition: "I share, therefore I am" [17] is the new credo, observed by clinical

psychologist Sherry Turkle. However, in our real, nonvirtual relationships, valuable and real human interactions are lacking [18].

Microblogging services such as Twitter, Reddit, Tumblr, and audiovisual platforms like Instagram have opened up new perspectives on presenting oneself and being seen by the digital, generalized Other. A recent development of *storytelling* via Instagram and Snapchat, where the user presents a short collection of photos and videos in a narrative context by adding captions to the visual material [19], emphasizes this short-lived, quick, and anecdotal view on an otherwise small detail of a user's life, as presented to one's followers.

Research on these online services has focused primarily on mental health aspects, with some studies suggesting that social networking sites can be addictive [20]. Problematic internet use and addictive behavior have been related to loneliness [21,22], but there is a lack of differentiated data on individual social media services. A survey among German students also showed positive correlations with mental health variables, which were interpreted as a consequence of the positive effects of sharing photos with a community. In contrast, the "interaction on Twitter seems to be more impersonal and less likely to enhance a person's social capital" [23] and thus correlates negatively with extraversion and self-esteem.

Despite a growing body of literature in this field, studies aimed explicitly at individualization are scarce and are mostly discussed within the context of narcissism [24], which is a psychopathological issue [25]. Many studies focus on the potentially harmful effects on young adults' mental health, and few include participants beyond the age of 30 years (such as in Chow and Wan's study on Facebook depression [26]). Age-specific results are needed in order to give personalized recommendations for preventive measures regarding the mental health of a cross-linked, social-media–using population.

The Quantified Self: Self-tracking and Self-optimization

Individualization has been reshaped recently due to a rising trend in the acquisition of personalized health information by wearable devices such as fitness trackers. This trend, which is often called *quantified-self movement* or *lifelogging*, is widespread, even beyond athletes. In 2016, the worldwide market reached up to 125 million devices, with estimates suggesting 237 million in 2020 [27]. Up to 69% of the US population regularly tracks at least 1 health indicator [28]. In Canada, a recent study found this number to be more than 66% [29]. The reasons for using self-tracking are, in most cases, increasing wellbeing or fitness, and sometimes, curiosity or the wish to question advice or a diagnosis delivered by a physician [30].

Individualization and Internet Usage

Studies on the self and social identity run into the danger of remaining theoretical at the expense of personal experiences [31]. The concept of generational change is one possible way out of this situation, using the context of societal generations coined by sociologist Karl Mannheim in the 1920s [32]. This perspective encourages a focus on the shared cultural imprinting

of a certain age group or cohort on the one hand and a differentiated look at single representatives of these groups at a psychological level on the other hand [33].

The Baby Boomer generation is characterized by "self-development, creativity, and pleasure of life" [34]. Regarding their usage of the internet at the beginning of social media development, Baby Boomers had the reputation of not understanding nonverbal cues in online communications such as emoticons [35]. However, Baby Boomers are now more familiar with social media and especially prefer forums and blogs for intellectual debates, entertainment, or sharing their own expertise [24].

The generation that followed the Baby Boomers, Generation X, is described as the "don't bother me" generation, which has withdrawn into private life [34] after the political turbulence of the 1960s, which were run by the generation before. Research considered this generation to be the "driving force behind" new media developments, especially the internet [36]. Nevertheless, research focusing on the social media use of this generation remains scarce [37].

Following the "X," the prominent Generation Y (also known as Millennials or Digital Natives, born between 1985 and 2000) are not committed to the values of former generations any longer, as they follow the mechanisms of "egotactics" [34]. This principle helps them "make flexible decisions in everyday life at any time. They use a mix of self-reference and a sensitive, strictly opportunistic, tactile, and tactical behavior, exploring opportunities and developmental potential. Ideals, norms, and principles are of little help here" [34]. This generation already grew up with the internet and does not remember a time before digitalization.

Many studies have described the changes caused by the internet and social media with respect to different age groups; however, how different generations experience the challenges and possibilities of the digital age, and whether there are any associations with different levels or degrees of individualization, has not yet been investigated.

Research Questions

This study aims to determine whether or not the younger generation is more concerned about individualization, with self-optimization at its strongest form, and if the internet and social media have further promoted this development. A secondary objective is to detect potential consequences of mental and emotional distress among different age groups due to the permanent availability of information about others that allows for social comparisons. We look for generational differences in the usage of the internet and social media that go along with the emotional experiences and expressions of self-relatedness.

Methods

Study Design and Recruitment

This qualitative study employed focus group discussions for data collection. Focus groups were originally developed in media research to measure the influence of a certain stimulus, a "focus," on an uninfluenced group [38]. This instrument has

increased in popularity, especially in health research [39]. The reason lies in the ease of getting access to the perspectives of participants. Especially, "not entirely encapsulated," nonreasonable responses can emerge in this unique form of communication [40].

We organized 6 focus groups, with at least 4 and a maximum of 8 participants from 3 consecutive generations of Baby Boomers, Generation X, and Digital Natives. We conducted 2 focus groups for each generation to collect adequate contributions for comparisons between the different age groups. We did not mix participants from different age groups since we wanted to facilitate a more relaxed and unbiased discussion flow [41]. All participants were of legal age.

Although the literature sets different limits for generations, we defined Baby Boomers as people born between 1950 and 1965, Generation X as those born between 1966 and 1980, and Digital Natives as those born between 1981 and 2000, following the definition of Palfrey and Gasser [42] in their famous publication, *Born Digital*.

The participants were recruited in Heidelberg University Hospital via public announcements, social networks (Facebook), and personal contact networks. They received a small monetary compensation for their participation.

Ethical approval for the study was granted by the Ethics Commission of the Medical Faculty of Heidelberg (S-039-2018) prior to recruitment. The Standards for Reporting Qualitative Research (SRQR) checklist was followed [43].

Procedure of the Focus Group Discussions

All focus group discussions followed a standardized procedure. First, the moderator presented 5 pictures on a flipchart titled "Me and My Devices." The pictures were of a smartphone, a tablet, a notebook, a desktop computer, and a smartwatch. The participants were asked to answer the question, "Which of these devices do you use regularly in your private, everyday life?" The first question had to be answered by every participant. The following questions could be discussed spontaneously in any order.

In the subsequent discussion, participants were initially asked about the importance they attached to the internet and mobile devices in their everyday life, how much time they spent with online media, and which platforms they usually visited. Further questions focused on communication, with whom and which groups it takes place, and if the usage of social media has changed communication and their respective social relationships in any way. The discussion flow was then directed to self-care and self-observation practices (ie, activities for self-tracking purposes, such as fitness apps, nutrition intake logs, or diaries with or without digital devices). The focus group guidelines are presented in Multimedia Appendix 1.

The first author (GM) moderated the focus groups, interfering only when the discussion lost its topic focus. The length of group discussions was between 76 and 82 minutes each.

Data Analysis

All discussions were transcribed manually and analyzed by using MAXQDA (version 12; Verbi GmbH) [44]. Coding was carried out by the first author (GM), a psychologist with a background in medical sociology and experience in research on eMental health, and was supported by advice from a colleague with a background in psychosomatics.

The analysis followed the methodology of inductive category formation in order to identify patterns of experiences and interpretations [45,46], whereby the categories are not built on theoretical considerations but directly from the data itself, which allows for an unprejudiced assessment. The first step was to find a deductive criterion of selection to build categories and to define the abstraction level "in a manner that fits best to the research question" [46]. We chose a low abstraction level to get the best approximate of the feelings and personal observations of the participants. The deductive criterion of selection focused on personal experiences of the internet and social media and every expression of self-relatedness, which could provide information about individualization. For example, self-tracking was considered to be a subcategory of self-relatedness. Multiple codings were possible in order to respect the complexity of single expressions. The categories were built up step by step while working through the whole transcription. As a rule of thumb, 10-30 categories were recommended, and 10%-50% through the text, the category system was built, redundant categories were summarized, and the material was reviewed.

All main codes, their definitions, subcodes, and supporting quotes are listed in Multimedia Appendix 2. In the text, we have provided the frequencies of main codes and subcodes of all participants in brackets behind the respective category in order to illustrate the main differences regarding the relevance of the respective code in comparison to other codes. To maintain textual readability, we have not presented the differences between the generations in the text; however, they are displayed in Multimedia Appendix 2. Descriptive statistics were calculated for the data on participants' characteristics and their internet-capable devices. We report means and standard deviations regarding the age of the sample. In all other data types, we present frequencies and percentages.

German Clinical Trials Register

This work is part of the study "Between self-care and self-optimization: The impact of the internet and social media on the identity of mentally burdened and unburdened people in the different generations," which was registered by the DRKS with ID DRKS00014815.

Results

Participant Characteristics

The sample consisted of 36 participants born between 1955 and 1998: 11 Baby Boomers, 10 members of Generation X, and 15 Digital Natives. The mean age for the Baby Boomers was 60.55 (SD 2.66) years, 47.40 (SD 3.13) years for the members of Generation X, and 23.80 (SD 3.19) years for the Digital Natives (Table 1).

Nearly all participants owned a smartphone, except for 1 member of Generation X. A desktop computer was owned by the majority of Baby Boomers, while the other generations were more likely to own a notebook or a tablet. Of the 36 participants, only 4 had a wristband: 1 in the group of Baby Boomers and 3 members of Generation X (Table 2).

Table 1. Baseline characteristics of the participants (N=36).

Participant Characteristics	Values
Age in years, mean (SD)	40.46 (16.33)
Digital Natives (born in 1988-1998)	23.80 (3.19)
Generation X (born in 1967-1976)	47.40 (3.13)
Baby Boomer (born in 1955-1962)	60.55 (2.66)
Gender, n (%)	
Male	17 (47.2)
Female	19 (52.8)

Table 2. Participant ownership of internet-capable devices (N=36).

Device	Baby Boomers, n (%)	Generation X, n (%)	Digital Natives, n (%)	Total, n (%)
Desktop computer	7 (64)	3 (30)	3 (20)	13 (100)
Notebook	5 (45)	8 (80)	14 (93)	27 (100)
Tablet	5 (45)	7 (70)	6 (40)	18 (100)
Smartphone	11 (100)	9 (90)	15 (100)	35 (100)
Wristband	1 (9)	3 (30)	0 (0)	4 (100)
Total number of devices owned by each generational group	29 (2.64)	30 (3.00)	38 (2.53)	97 (2.69)

Focus Group Discussions

By means of inductive category formation, 5 main themes of information were found: (1) reasons for the use of the internet and social media, (2) effects of the use of the internet and social media, (3) personal feelings and experiences, (4) self-relatedness, and (5) social relationships.

Reasons for the Use of the Internet and Social Media

During the course of the discussions, all participant statements that provided insight into the causes and occasions of internet and social media use in daily life were coded with the main code, *Reasons for the use of the internet and social media.* This category was found in 329 cases. The most important reasons were communication (122/329), organization of daily affairs (55/329), and information search (53/329). Other causes related to job or study (45/329), entertainment (41/329), creativity (9/329), and pastime (4/329).

Baby Boomers put the most emphasis on communication, job/study, and information search. Entertainment was a minor issue for them, and while they organized their daily affairs online, they did not do so as often as the younger generations. A typical quote of one individual from the Baby Boomer group was,

> *In a conversation, I can ask Dr. Google or Wikipedia. At our last family reunion, I was very proud. Our city guide … said something, that and that. And I was so sure that this couldn't be right, and then I checked, and it was me who was right!* [Baby Boomer, male, 64, group BB1]

In contrast, Generation X reported the importance of the online organization of daily affairs even more often than communication, followed by information search. Job/study and entertainment was important to them to a certain extent, and this generation reported somewhat creative activities as a reason to be online. A typical quote of a Gen X member was,

> *I don't have an alarm clock either. Only the smartphone does that. Of course, I don't write letters anymore either… what for? There is mail for that. And WhatsApp groups are very, very important, also to keep in touch with the family. And, of course, apps, so there are apps like "sand by the sea." No matter if … if you … so if I want to cook something, I get a recipe via app …* [Generation X, male, 48, group X1]

This quote reflects the subcode "entertainment," which was a relevant issue to this generation.

> *… what I like to do on the internet is watching YouTube videos. Different fail-videos from time to time, where people fall down and so, sometimes that is quite funny.* [Generation X, male, group X1]

The youngest generation, the Digital Natives, put the strongest focus on communication in comparison with the other groups, followed by entertainment and the organization of daily affairs. As well, they often look for quick information and use the internet for their study. In some cases, they use their mobile phones for mere pastime.

> *Well, I also use Instagram, but more when I'm bored or when I think, I've got about 15 minutes left before I have to leave the house and it's not worth the effort to start something. Then I scroll down a bit …* [Digital Native, female, 20, group DN1]

Effects of the Use of the Internet and Social Media

The code *Effects of the use of the internet and social media* summarized discussion contributions that referred to the observed effects of the use of the internet and social media on the participants' daily lives or on society. In total, 501 codings were found. The most important issues in this category were described by comparison of generations (100/501), followed by communication (92/501) and societal changes (77/501). Next, relevant aspects were simplification (54/501), outdated technology (eg, VHS; 43/501), language (27/501), and writing letters or postcards (25/501). Minor topics the participants talked about were financial issues (23/501), telephoning (20/501), loss of abilities (15/501), health (12/501), help/sharing (9/501), and environment (4/501).

The Baby Boomer groups placed emphasis on the comparison of generations, societal changes, and changes in communicative behavior. Further, they discussed outdated, formerly "new" technologies like VHS tapes, and they talked about writing letters or postcards, which they fear will become obsolete as well. They often observe changes in language.

> *In the WhatsApp group …sometimes I get the impression that I am the only one in my family who still knows something like punctuation.* [Baby Boomer, 62, male, group BB1]

Members of Generation X attached the highest importance to how new technologies simplified their daily affairs. In this context, they draw many comparisons between generations, talking about changes in communication and language. At the same time, they see a potential loss of abilities. Finally, financial issues seem important to them.

> *I also have my Payback card in my cell phone. I no longer have a plastic card. Download the app, and then you have a bar code and put it on their device, and then you have your Payback card.* [Generation X, female, 51, group X1]

While talking about the effects of the internet and social media, the Digital Natives had a tight focus on communication, societal changes, and comparisons of generations. They often talked about staying connected with members of the Baby Boomer generation, who were their parent generation. Further aspects were financial issues, health topics, and telephoning. Typical quotes regarding changes in communication referred to social relationships.

> *I think that maybe WhatsApp and etcetera have made friendships a bit deeper, that you just get a lot more from each other.* [Digital Native, female, group DN1]

Another quote illustrates the subcoding Comparison of generations in the context of communication as well.

> *I think it's also a completely different writing style, how we write among each other and how I write with*

my parents. So, my mother just writes with dots and capitalization and so on. We really do not have that anymore. [Digital Native, female, 20 years, group Y1]

Personal Feelings and Experiences

The code *Personal feelings and experiences* were given in statements that expressed intimate observations of the participants regarding their emotions after or during the usage of the internet and social media. All groups expressed this in various forms, with 381 codings in total. The predominating subcategory was challenges (173/381), further differentiated into solitude, the feeling of permanent availability, norms and expectations felt by others, dissatisfaction, liability, time pressure, and acceleration. The second important domain of feelings were fears (123), which were expressed as fear of commercial interests, fear of surveillance, distrust and uncertainty, and feeling overstrained. Positive emotions (47) were expressed by talking about feeling anonymous, not having to go back to being without the internet, enjoying offline time, and curiosity and fascination. Some group members felt risks (31), especially the risk of addiction or loss of reality. Additionally, the participants showed indifferent emotions (7) (eg, by dismissing potential surveillance as unimportant). Further details are presented in Multimedia Appendix 2.

The feelings of the Baby Boomer generation were dominated by fears. Mistrust or feeling insecure regarding the quality of information provided by the internet and the fear of potential surveillance due to a lack of data security were expressed in many cases. Feeling overstrained and the fears of commercial interests were both topics of discussion.

The second important group of feelings related to challenges; the Baby Boomers often felt pressure from feelings of liability and being forced to stay permanently available with their mobile devices. Indifference was not shown at all. Positive emotions and risks were both present, but to a small degree.

> *I think something is slipping away, that's the danger … What actually happens to the data that I have entered there? Where do they end up? And what might be done with it?* [Baby Boomer, female, 61, group BB2]

Fears and challenges were expressed in the Generation X group as well. If they talked about fears, their concerns related to potential surveillance and mistrust or insecurity. Potential risks were felt by losing contact with reality. Positive emotions played the most important role in comparison to the other age groups. If positive, their emotions were described by the subcode curiosity and fascination.

> *Yes, but I find it exciting. I think that's interesting. And there you can make a data comparison, women my age, menopause and so on. You can just watch everything a bit. I think that's exciting.* [Generation X, female, 51, group X1]

The Digital Natives put the most emphasis on challenges that they experienced in the usage of the internet and social media. They felt pressured by the norms and expectations of others, such as the necessity to be online, to quickly respond to a message, and to be available all the time. If they expressed fears, these were related to potential surveillance, but in this group, some members voiced feelings of indifference at the same time. They accepted the fact of surveillance because they were aware that, otherwise, most functions would be useless if the terms of use of, for example, Facebook, were rejected. Positive emotions related to curiosity and fascination. Some members reported the feeling of risks regarding addiction or loss of reality. A typical example regarding norms and expectations are reflected in the following quote:

> *Yes, because it is such a social filter. Because everyone knows, he knows the rules of Instagram. I know I have to jazz everything up and somehow say, "Wow, it's so cool what you posted there." So this is such a love culture there, isn't it? Where to exaggerate …* [Digital Native, male, 28, group DN2]

Self-relatedness

The main category, *Self-relatedness,* included expressions that showed an orientation towards person-based and individual choices. This code was used in all cases in which the behavior with the internet and social media was related to the individual (eg, to the surveillance of personal health, to self-presentation in a network or forum, or to personal organizing like calendars or diaries). In sum, 330 codings were found. Subcodes were distancing (114/330; ie, the conscious decision against being permanently available or following a specific internet trend), self-tracking (62/330), self-reflection (44/330), self-control (39/330), self-optimization (22/330), autonomy (11/330), and self-presentation (9/330). These aspects all referred to one's own person. Self-relatedness observed in others (29/330) was also found in this domain.

The Baby Boomers expressed the least self-relatedness in comparison to the other generations, and they hardly talked about individual, person-based issues. They often expressed the wish to distance themselves from several developments of the internet and social media that they consider to be useless or even dangerous to them. Attempts of self-control refer to self-organization by using calendar apps or messaging to oneself, and self-reflection was sometimes done in an ironic way in the course of distancing from new technologies (eg, "I'm not ready for that yet"). If self-tracking was present, it was used for health care reasons. Self-optimization was no issue at all, nor was self-presentation in most cases.

> *I had a good feeling, for I could say, ''I won't join to do that'' (regarding Facebook).* [Baby Boomer, male, 64, group BB2]

Regarding expressions of self-relatedness, Generation X also talked a lot about distancing, especially as far as self-presentation in platforms was concerned. Self-tracking was often done for several reasons; however, self-optimization was mostly not an issue, nor was the presentation of one's own self in networks a matter of discussion. Self-reflection, seeking autonomy, and self-control measures were minor topics that referred to calendar apps, date reminders and notifications for self-organization, etc.

Yes, so I go running regularly. Do the Runtastic, which is also an app. But I have to say, what bothered me again, massively, that I could upload it to Facebook by pushing a button. So that everyone can see it, too. I thought that was stupid again. [Generation X, male, 47, group X1]

Self-presentation was observed in others in a critical way.

It is a self-presentation. I was here on vacation, I have this, I have this, I have that, I was always eating, I have a lot of friends... [Generation X, female, 48, group X1]

To the youngest generation, self-tracking was the most important subcategory in the context of self-relatedness. In tracking their behavior, sports activities or nutrition tracking were done, in many cases, explicitly for reasons of self-optimization. As well, the Digital Natives reported attempts to distance themselves from single forms of media usage; however, these attempts often fail. Self-reflection and self-presentation were a frequent issue of discussion as well.

I often notice that I am just grabbing it (the smartphone) relatively unconsciously and looking at something, ... if you are walking alone through the city, in order to seem engaged ... just take it and pretend to continue your education or something (laugh). That's quite ... a bit scary. [Digital Native, male, 21, group DN2]

Another example shows the subcategory of self-relatedness observed in others, which was another common topic of discussion among the Digital Natives.

Many do this to compare themselves with friends. For example, I also know one who always posts that on Instagram, that she is the best, and that she walked most of the time or so (laugh) ... I think that's a bit silly. Posting such things is really just pushing yourself. [Digital Native, female, 22, group DN2]

Social Relationships

Statements that referred to online communication with others via the internet and social media (eg, with friends, colleagues, or family members) were titled *Social relationships*. In all generations, 214 codings referred to the main category *Social relationships*. The most often used subcode referred to social comparisons (70/214), followed by the relationship of the participants to their partners/friends (68/214) and family (53/214). Minor issues discussed in the groups were conflicts (8/214), gender stereotypes (6/214), associations/engagement (5/214), and social inequities (4/214).

The Baby Boomers described maintaining close relationships with their children or grandchildren with the help of their mobile devices. Staying connected to partners/friends was important to them as well. In this generation, few participants were members of the specific associations with which they were engaged. Social comparisons and conflicts were mostly not an issue. Gender stereotypes and social inequities were not discussed.

I'm in a WhatsApp group that only affects my ... core family, and since my 2 adult children have moved out, that's important, too, to maintain some social contact. [Baby Boomer, male, 62, group BB1]

Members of the Generation X group cared for their relationships, especially with partners and friends. Social comparisons were often made. Only then, in third place, staying connected to family was a matter of discussion. Gender stereotypes and social inequities were not discussed, and conflicts and memberships in associations and respective engagement were discussed to a small degree. The following example illustrates the kind of contact a Gen X woman described with her friends:

There's a lack of personal contact; it has become less. In the past, you just took the phone or you met somebody, and today you quickly write a message. If it is someone's birthday, I avoid sending only a WhatsApp message; I actually call and congratulate. [Generation X, female, 44, group X2]

The Digital Natives often addressed social comparisons directly and described their relationship to friends and partners as being shaped by online activities. In some cases, the Digital Natives talked openly about conflicts that emerged from misunderstandings due to social media messages. As well, potential gender stereotypes and social inequalities that manifest on Instagram were discussed.

For me personally, I find social comparisons much more difficult if they are people I know and ... if they post a picture every day, but I have to stay in the library, I think to myself, Okay, why do they manage that and I don't? [Generation Y, male, 28 years, group DN2]

Discussion

Principal Findings

This study aimed to investigate the thesis of an ongoing individualization process [1,2,4] and how the internet and social media have contributed to this development. We looked for generational differences in the usage of the internet and social media that go along with emotional experiences and expressions of self-relatedness. We analyzed 6 age-specific focus group discussions, with 3 consecutive generations of Baby Boomers (born in 1950-1965), Generation X (born in 1966-1980), and Digital Natives (born in 1981-2000). If the ongoing individualization process is promoted by the use of the internet and social media, data should reveal that the generation of Digital Natives uses the internet in a more individualized, self-presenting, and self-optimizing way in comparison to older age groups. Moreover, if growing individualization can be attributed to the use of the internet and social media, we investigated which potential emotional and mental health consequences might be traced back to this development. Given the fact that psychological strain due to individualization cannot be observed directly, we regarded this issue by analyzing the expressions of feelings and experiences in the context of statements of self-relatedness.

Our findings suggest that the Digital Natives expressed more tendencies toward self-relatedness than the Baby Boomers, but less than the next recent generation, Generation X. The usage of self-tracking technology for fitness and health parameters was obvious in both younger generations. However, these 2 generations explained their choice for tracking technologies in a substantially different way. Generation X showed curiosity and fascination for the possibility of tracking personalized data, but not desires for self-optimization. In contrast, the Digital Natives reported making use of calorie-tracking apps or fitness apps with the aim of optimizing their own bodies. A previous study had investigated predictors of fitness and weight loss websites and found a close relationship between the internalization of beauty ideals, being female, and a high frequency of internet use in a population with a median age of 24 years [47].

Social networking played an important role for both younger generations but, again, for different reasons. While Generation X presented themselves as critical about posting vacation pictures or private information online, they were often members of Facebook or other social networking sites. Meanwhile, the Digital Natives were more inclined to present themselves regularly on microblogging-services like Instagram. They felt the need to stay visible and often reported observing this behavior in others.

Among Baby Boomers, the self-presentation of younger generations on social networking sites was observed with a sense of alienation. This generation expressed more reluctance and was not used to talking about personal issues in a group. In contrast, the Digital Natives talked openly about their personal entertainment, their fitness, and forthcoming events. This does not necessarily mean that the elders did not care about themselves, as this behavior might just be a communicative style. In another study carried out in a different cultural and communicative context (Hong Kong), similar generations were investigated regarding their social media use, and many facets of narcissism were found amongst Baby Boomers [24]. Future research could integrate both concepts: individualization as a kind of self-care on the one end of a potential continuum, and narcissism on the other end.

Generation Profiles

As generations cannot be characterized entirely without expressions of self-relatedness, generation profiles regarding the characteristics of internet and social media use in each age group were used to contribute to a deeper understanding of the relationship between internet usage and individualization for each group. Also, the possible mental and emotional strains that may lead to a mental disorder must be reflected in each generation separately.

Baby Boomers

The Baby Boomers presented themselves as open-minded but were critical about their use of the internet and social media. Many of them used the internet as a "paperless newspaper" with a desktop computer or a smartphone, which was switched on and off at a certain time of the day. They appreciated the possibility of staying connected with their children and grandchildren but, at the same time, expressed feelings of distrust regarding the provenience of information or being overwhelmed by technical requirements. Baby Boomers are sometimes called "digital immigrants" [48] to describe the fact that many Baby Boomers have still not mastered computerization; for example, the use of emoticons is not common to every member of this generation [35]. Research on self-tracking of health-related data points in a similar direction: not all self-tracking is digital self-tracking. A substantial part of the self-trackers are "traditional" trackers (ie, do tracking manually by writing medical data on a sheet or in a booklet). While digital trackers are mainly men, young, and of higher income, the traditional trackers are often women, older than 55 years, and have lower income or are retired [29].

In our study, this generation aimed at autonomy and self-determination to a certain extent. Even when distancing themselves from certain unappreciated internet trends, they were concerned about their own development, often making person-based choices against the use of the newest trendy platform or messaging service. They wanted to stay connected to their own network and family members. Our result goes along with an earlier investigation of generational differences in social media networking virtual needs [48], which highlighted that the Baby Boomers especially, but also Generation X, show needs for competency and a sense of being effective. In regard to possible mental and emotional strains, our results show that while the Baby Boomers have to cope with the need to stay updated and feel the pressure of social-environmental expectations, they lack a deeper understanding of the online activities of their children and grandchildren. Psychosomatic research supports this observation: If still employed, Baby Boomers feel a high pressure in competing with younger colleagues [49]. Targeted coping strategies are needed for this generation, regardless of whether they are still employed or not.

Generation X

Our Generation X participants showed a great variety of internet and social media usage. This generation possessed the highest number of internet-capable devices per person, followed by the Baby Boomers and Digital Natives. Some members of this generation were technophiles (ie, curious to try wearable devices and used many apps to simplify their daily affairs). Consequently, the possibility of data surveillance was accepted as an unavoidable circumstance. This observation that Generation X is interested in forming their own identity in a positive, open-minded way was made by a cross-cultural study among German, Japanese, and US participants [36]. In our study, some of the Generation X group members appreciated personalized advertising, which is normally discussed as a harmful form of surveillance [50]. This reveals an inherent desire for personalized services geared toward the accommodation of individual wishes. Some participants presented themselves as rather restrictive due to fears of data surveillance. Nonetheless, they expressed this in a highly self-focused way by talking about concerns about their own health or regarding various risks, like the commercial interests of app developers.

Generation X reported being especially active on Facebook. Numerous studies have pointed to mental health, personality factors, and narcissism as predictors for the excessive use of Facebook. Several diverging results regarding the usage of Facebook show that a simple "Facebook is harmful for mental health" statement is not possible. A meta-analysis found associations between the use of social networking sites like Facebook and depression. The authors show that this association can be explained by the users' social comparisons and not by the mere time spent on the sites [51], which hints at the same underlying mechanisms found in our study. Other studies reveal that the correlation of depressive symptoms with the excessive use of Facebook can be explained with high degrees of neuroticism [26]. Facebook users, in comparison to nonusers, tend to show more narcissism and self-esteem, as well as higher degrees of life-satisfaction [23]. Resilience factors like self-esteem and curiosity should be strengthened in this generation while taking data surveillance fears seriously.

Digital Natives

The youngest generation, the Digital Natives, were not always as self-focused as they are reputed to be in a previous German survey study [34]. Satisfied with owning a smartphone alone and sometimes a notebook, they expressed intentions of adequately communicating with their parents or grandparents' generation. Additionally, they showed a high level of reflection about their own behavior on the internet and social media. They were critical observers of others' self-presentation, especially regarding the use of platforms like Instagram or Snapchat.

It is noteworthy that the youngest generation felt significantly more pressured by the norms and expectations of their social environment (eg, by having to answer immediately to a text message, staying on Instagram every few hours, or posting personal updates regularly). Furthermore, they drew many social comparisons based on information delivered by the internet ("My colleagues are having fun; I have to study, and this means that there is something wrong with me"). Not surprisingly, "enjoying offline time" was a particular matter of discussion in this generation.

Future Implications

Social comparisons are a natural part of social exchange and play an important role in identity formation and thus are a necessary step in growing up and reaching adulthood [9]. The Digital Natives in our study were adults but may still be undergoing identity formation, which, according to the renowned psycho-analytical researcher, Erikson [11], is a life-long process. The question arises, then, as to whether the comparisons felt by Digital Natives are a precondition or a result of their identity formation. Ultimately, the immediacy of information about the social activities delivered via social media has the potential to cause a continuous, subtle strain on the identity of the younger generation, which is cross-linked by different types of technology. This does not need to end up in dramatic cyberbullying scenarios, which are often reported among adolescents [52,53], but may lead to increased mental distress for people vulnerable to mental health problems.

Studies that concentrate on patients with mental disorders show contradictory results. Patients with mental health disorders feel negatively affected by social media applications but believe that health care apps and calendars can positively influence their mental health [54]. However, patients with severe mental illnesses reportedly benefited from social media usage, which helped them participate in social life [55]. New methodological instruments like the social media disorder scale have emerged and will provide a more differentiated view on the topic [56].

Previous research on mental health and social media tended to focus on the apparent and strong effects of internet addiction due to extensive gaming [21] or eating disorders [57]. There is a strong need to promote age-specific investigations on more subtle associations, like performance pressure due to social comparisons among younger generations, or feelings of mistrust and overwhelm among older generations. The categories found in this study might serve as a template for the development of age-specific questionnaires after integrating further results of other cultures. This study design could be part of an exploratory design, which normally starts with an in-depth qualitative analysis and leads to a broader, epidemiologic perspective [58].

Limitations

The definition for age groups of specific generations varies significantly in the literature. Hurrelmann [34], for example, defines a generation in 15-year periods, with Generation X starting in 1970 and ending in 1985. While other studies agree that this generation started in 1965, the ending year ranges from 1976 [24], 1978 [36], 1982 [33], and even to 1983 [48]. However, we used the definitions provided by Palfrey and Gasser [34] regarding Digital Natives and went back in increments of 15 years, which may or may not be the universally accepted age groups.

This study followed all principles of good scientific practice; however, the most challenging limitation was the issue of age effects versus cohort effects, which is a recognized problem in social sciences. As Schröder [59] remarked, it is not possible to carve out differences between generations without controlling for age as a variable. This problem might only be resolved by choosing a longitudinal study design.

Moreover, the qualitative coding and interpretative work were done by the author alone, so no interrater reliability can be provided. In addition, our participants were recruited in a predominantly academic environment at the University of Heidelberg in Germany; it is understood that these people may use the internet and their social networks in more ambitious ways than might be generalizable to people of other professions or countries.

Conclusions

The Digital Natives and Generation X participants expressed more individualistic behaviors, like self-tracking, self-optimization, and self-presenting in social networks, than Baby Boomers. They also often observed self-presenting behavior in friends and colleagues. In contrast, the Baby Boomers were less driven by individualization but seemed very interested in taking advantage of the possibility of staying informed and connected while not following every new

technological trend. They kept a critical distance from the internet and social media but also did not want to be left behind the developments.

The most striking difference among the generations was the high pressure of norms and expectations of the social environment and the social comparisons felt by the Digital Natives. It remains unclear if this is a sign or a precondition of individualization, but it is known that social comparisons may result in mental distress. Further investigations on this association are necessary to promote a mentally healthy way of using the internet and social media.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the main categories and subcategories used for the focus groups.
[DOCX File , 16 KB - jmir_v22i11e20528_app1.docx]

Multimedia Appendix 2
Inductive codes and subcodes in the generations (total frequencies).
[DOCX File , 18 KB - jmir_v22i11e20528_app2.docx]

References

1. Giddens A. Modernity and self-identity. Cambridge: Polity Pr; 1991.
2. Beck U. Risikogesellschaft: auf dem Weg in eine andere Moderne. Frankfurt a. Main: Suhrkamp; 2016.
3. Hofmeister H. Individualisation of the life course. International Social Science Journal 2015 May 13;64(213-214):279-290. [doi: 10.1111/issj.12053]
4. Elias N. Die Gesellschaft der Individuen. Frankfurt (Main): Suhrkamp; 2001.
5. Mills M. Individualization and the life course: Toward a theoretical model and empirical evidence. In: Howard C, editor. Contested individualization. New York: Palgrave Macmillan US; 2007:61-79.
6. Beck U, Beck-Gernsheim E. Losing the traditional: Individualization and "precarious freedoms". In: Elliott A, du Gay P, editors. Identity in question. London: SAGE publications; 2009:13-36.
7. Ehrenberg A. Das erschöpfte Selbst: Depression und Gesellschaft in der Gegenwart. Frankfurt a.M: Campus Verlag; 2015.
8. Côté JE, Schwartz SJ. Comparing psychological and sociological approaches to identity: identity status, identity capital, and the individualization process. Journal of Adolescence 2002 Dec;25(6):571-586. [doi: 10.1006/jado.2002.0511]
9. Noack J. Erik H. Erikson: Identität und Lebenszyklus. In: Jörissen B, Zirfas J, editors. Schlüsselwerke der Identitätsforschung. Wiesbaden: VS Verlag für Sozialwissenschaften; 2010:37-53.
10. Fuchs-Heinritz W, Lautmann R, Rammstedt O, Wienold H. Lexikon zur Soziologie. Wiesbaden: Westdeutscher Verlag GmbH, Opladen; 1994.
11. Abels H. Identität – sich selbst zum Objekt machen. In: Abels H, editor. Identität: Über die Entstehung des Gedankens, dass der Mensch ein Individuum ist, den nicht leicht zu verwirklichenden Anspruch auf Individualität und Kompetenzen, Identität in einer riskanten Moderne zu finden und zu wahren. Wiesbaden: Springer Fachmedien; 2017:203-218.
12. Goffman E. Stigma: über Techniken der Bewältigung beschädigter Identität. Frankfurt am Main: Suhrkamp; 2010.
13. Mead GH. In: Morris CW, editor. Geist, Identität und Gesellschaft aus der Sicht des Sozialbehaviorismus. Frankfurt am Main: Suhrkamp; 2013.
14. King V, Gerisch B, Rosa H, Schreiber J, Salfeld B. Überforderung als neue Normalität. In: Fuchs T, Iwer L, Micali S, editors. Das überforderte Subjekt: Zeitdiagnosen einer beschleunigten Gesellschaft. Berlin: Suhrkamp Verlag; 2018.
15. Borkenhagen A, Brähler E. Die Selbstverbesserung des Menschen: Wunschmedizin und Enhancement aus medizinpsychologischer Perspektive. Gießen: Psychosozial Verlag; 2012.
16. Rosa H. Beschleunigung: die Veränderung der Zeitstrukturen in der Moderne. Frankfurt/Main: Suhrkamp; 2005.
17. Turkle S. Connected, but alone).: TED Conferences; 2012. URL: https://www.ted.com/talks/sherry_turkle_alone_together?language=de [accessed 2020-08-01]
18. Turkle S. Verloren unter 100 Freunden: wie wir in der digitalen Welt seelisch verkümmern. München: Riemann; 2012.

19. Oswald B. Instagram-Storys: Die kreative Art des Geschichtenerzählens. 2019. URL: https://www.fachjournalist.de/instagram-storys-die-kreative-art-des-geschichtenerzaehlens/ [accessed 2020-08-01]

20. Kircaburun K, Griffiths MD. Instagram addiction and the Big Five of personality: The mediating role of self-liking. Journal of Behavioral Addictions 2018 Mar;7(1):158-170. [doi: 10.1556/2006.7.2018.15]

21. Costa RM, Patrão I, Machado M. Problematic internet use and feelings of loneliness. International Journal of Psychiatry in Clinical Practice 2018 Dec 20;23(2):160-162. [doi: 10.1080/13651501.2018.1539180]

22. Savolainen I, Oksanen A, Kaakinen M, Sirola A, Paek HJ. The Role of Perceived Loneliness in Youth Addictive Behaviors: Cross-National Survey Study. JMIR Ment Health 2020 Jan 2;7(1):e14035. [doi: 10.2196/14035]

23. Brailovskaia J, Margraf J. What does media use reveal about personality and mental health? An exploratory investigation among German students. PLoS ONE 2018 Jan 25;13(1):e0191810. [doi: 10.1371/journal.pone.0191810]

24. Leung L. Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. Computers in Human Behavior 2013 May;29(3):997-1006. [doi: 10.1016/j.chb.2012.12.028]

25. Brailovskaia J, Bierhoff HW. Generation Me: Soziale Medien und Narzissmus. PiD - Psychotherapie im Dialog 2019 Sep 03;20(03):26-30. [doi: 10.1055/a-0771-7047]

26. Chow TS, Wan HY. Is there any "Facebook Depression"? Exploring the moderating roles of neuroticism, Facebook social comparison and envy. Personality and Individual Differences 2017 Dec;119:277-282. [doi: 10.1016/j.paid.2017.07.032]

27. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in Medicine. Adv. Mater 2018 Jun 11;30(33):1706910. [doi: 10.1002/adma.201706910]

28. Hilty DM, Chan S, Hwang T, Wong A, Bauer AM. Advances in mobile mental health: opportunities and implications for the spectrum of e-mental health services. mHealth 2017 Aug 21;3:34-34. [doi: 10.21037/mhealth.2017.06.02]

29. Paré G, Leaver C, Bourget C. Diffusion of the Digital Health Self-Tracking Movement in Canada: Results of a National Survey. J Med Internet Res 2018 May 02;20(5):e177. [doi: 10.2196/jmir.9388]

30. Neff G, Nafus D. Self-tracking. Cambridge, Massachusetts: The MIT Press; 2016.

31. Callero PL. The Sociology of the Self. Annu. Rev. Sociol 2003 Aug;29(1):115-133. [doi: 10.1146/annurev.soc.29.010202.100057]

32. Steinbach A. Generation. In: Kopp J, Steinbach A, editors. Grundbegriffe der Soziologie. Wiesbaden: Springer Fachmedien; 2018:129-131.

33. Haller M. Was wollt ihr eigentlich? Die schöne neue Welt der Generation Y. Hamburg: Murmann Publishers GmbH; 2015.

34. Hurrelmann K. In: Albrecht E, editor. Die heimlichen Revolutionäre: Wie die Generation Y unsere Welt verändert. Weinheim: Beltz; 2014.

35. Krohn FB. A Generational Approach to Using Emoticons as Nonverbal Communication. Journal of Technical Writing and Communication 2016 Aug;34(4):321-328. [doi: 10.2190/9eqh-de81-cwg1-qll9]

36. Hachtmann F. Generation X revisited: An exploratory cross-cultural case study. : American Academy of Advertising; 2008 Presented at: Annual Proceedings American Academy of Advertising; 27.-30.03.2008; San Mateo p. A.

37. Hartijasti Y. Exploring the motivation in using Facebook: A comparative study between Generation XY in Indonesia. 2013;20(1):53-66.

38. Przyborski A, Riegler J. Gruppendiskussion und Fokusgruppe. In: Mey G, Mruck K, editors. Handbuch qualitative Forschung in der Psychologie. Wiesbaden: Springer; 2010:436-448.

39. Wong LP. Focus group discussion: a tool for healthmedical research. Singapore Med J 2008;49(3):256-260.

40. Kitzinger J. Qualitative Research: Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302. [doi: 10.1136/bmj.311.7000.299]

41. Zwick M, Schröter R. Konzeption und Durchführung von Fokusgruppen am Beispiel des BMBF-Projekts "Übergewicht und Adipositas bei Kindern, Jugendlichen und jungen Erwachsenen als systemisches Risiko?". In: Schulz M, Mack B, Renn R, editors. Fokusgruppen in der empirischen Sozialwissenschaft. Wiesbaden: VS Verlag für Sozialwissenschaften; 2012:24-48.

42. Palfrey J, Gasser U. Born digital: Understanding the first generation of digital natives. New York: Basic Books; 2008.

43. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research. Academic Medicine 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]

44. VERBI. MAXQDA, Software für qualitative Datenanalyse, 1989 – 2020. Berlin: VERBI Software. Consult. Sozialforschung GmbH; 2016. URL: https://www.maxqda.de/ [accessed 2020-08-01]

45. Mayring P. Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken. Weinheim, Basel: Beltz; 2016.

46. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. Weinheim, Basel: Beltz; 2015.

47. Almenara CA, Machackova H, Smahel D. Sociodemographic, Attitudinal, and Behavioral Correlates of Using Nutrition, Weight Loss, and Fitness Websites: An Online Survey. J Med Internet Res 2019 Apr 04;21(4):e10189. [doi: 10.2196/10189]

48. Krishen AS, Berezan O, Agarwal S, Kachroo P. The generation of virtual needs: Recipes for satisfaction in social media networking. Journal of Business Research 2016 Nov;69(11):5248-5254. [doi: 10.1016/j.jbusres.2016.04.120]

49. Tophoven S, Tisch A, Rauch A, Burghardt A. Psychosoziale Arbeitsbedingungen und mentale Gesundheit der Babyboomergeneration. Gesundheitswesen 2015 Mar 11;77(04):e63-e69. [doi: 10.1055/s-0034-1398596]

XSL•FO

RenderX

50.   Fuchs C. New Media, Web 2.0 and Surveillance. Sociology Compass 2011;5(2):134-147. [doi: 10.1111/j.1751-9020.2010.00354.x]

51.   Yoon S, Kleinman M, Mertz J, Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook–depression relations. Journal of Affective Disorders 2019 Apr 01;248:65-72. [doi: 10.1016/j.jad.2019.01.026]

52.   Modecki KL, Minchin J, Harbaugh AG, Guerra NG, Runions KC. Bullying Prevalence Across Contexts: A Meta-analysis Measuring Cyber and Traditional Bullying. Journal of Adolescent Health 2014 Nov;55(5):602-611. [doi: 10.1016/j.jadohealth.2014.06.007]

53.   Betts LR, Spenser KA, Gardner SE. Adolescents' Involvement in Cyber Bullying and Perceptions of School: The Importance of Perceived Peer Acceptance for Female Adolescents. Sex Roles 2017 Mar 15;77(7-8):471-481. [doi: 10.1007/s11199-017-0742-2]

54.   Beard C, Silverman AL, Forgeard M, Wilmer MT, Torous J, Bjo¨rgvinsson T. Smartphone, Social Media, and Mental Health App Use in an Acute Transdiagnostic Psychiatric Sample. JMIR Mhealth Uhealth 2019 Jun 7;7(6):e13364. [doi: 10.2196/13364]

55.   Brusilovskiy E, Townley G, Snethen G, Salzer MS. Social media use, community participation and psychological well-being among individuals with serious mental illnesses. Computers in Human Behavior 2016 Dec;65:232-240. [doi: 10.1016/j.chb.2016.08.036]

56.   van den Eijnden RJ, Lemmens JS, Valkenburg PM. The Social Media Disorder Scale. Computers in Human Behavior 2016 Aug;61:478-487. [doi: 10.1016/j.chb.2016.03.038]

57.   Turner PG, Lefevre CE. Instagram use is linked to increased symptoms of orthorexia nervosa. Eat Weight Disord 2017 Mar 1;22(2):277-284. [doi: 10.1007/s40519-017-0364-2]

58.   Creswell JW. A concise introduction to mixed methods research. Thousand Oaks: SAGE publications; 2014.

59.   Schröder M. Der Generationenmythos. Kölner Zeitschrift für Soziologie und Sozialpsychologie 2018 Oct 2;70(3):469-494. [doi: 10.1007/s11577-018-0570-6]

Original Paper

Factors Related to the Behavior of People Who Have Never Used the Internet for Voluntary Reasons: Cross-Sectional Survey Study

Keeho Park[1], MD, PhD

Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Republic of Korea

Corresponding Author:
Keeho Park, MD, PhD
Graduate School of Cancer Science and Policy
National Cancer Center
323 Ilsan-ro Ilsandong-gu
Goyang, 410769
Republic of Korea
Phone: 82 9202160
Fax: 82 9202929
Email: park.keeho@gmail.com

Abstract

Background: If there are people who do not want to use the internet despite having the circumstances and conditions for using it, another policy consideration will be needed.

Objective: The purpose of this study was to explore the factors related to the behavior of people who do not voluntarily use the internet.

Methods: A cross-sectional survey was conducted in 2018. It used a proportional quota random sampling design to select a representative sample of Koreans. Accordingly, 6150 participants were included in the study. Multiple logistic regression methods were used to explore the predicting factors of the act of voluntarily not using the internet.

Results: Age, education level, bonding and bridging social capitals, and daily life satisfaction for health status were found to be factors related to the behavior of not voluntarily using the internet. However, gender, household income, occupation, family size, and community type were not related to voluntary nonuse of the internet.

Conclusions: It was found that sociodemographic factors, such as age and education level, which are difficult to modify, along with psychosocial factors located deeper than the visible living conditions, such as social capital and life satisfaction, are involved in voluntary internet nonuse. These results also suggest that it is not desirable to proceed with policies related to information and communications technology on a separate track, but rather that they should be comprehensively approached with other social policies that design various social interventions in order to enhance equity within the society.

KEYWORDS

internet; use; voluntary; factor; information and communications technology; digital divide

Introduction

The internet is now an essential part of life. However, this may not be the case in countries where internet infrastructure is not available nationwide or in countries where there are few people who can pay for internet services. Even if people do not live in such a country, there may be variations in the available digital life within the country or community from a digital divide caused by various factors.

In countries with advanced information and communications technology (ICT), attempts are underway to actively utilize these technologies for providing social services [1]. Notably, even if internet access is improved for those who are targeted for social welfare services, such efforts will not be effective unless the fundamental social vulnerabilities that interfere with the intention to use the internet are addressed.

When it comes to the internet, its use is very common in South Korea, where there is a good environment that allows it to be effective. South Korea is the 12th largest country in the world according to gross domestic product (GDP) [2]. As of 2018, the internet access rate was the highest in South Korea at 99.5% (including wired and wireless), followed by Iceland (99.2%),

the Netherlands (98.0%), Norway (96.0%), the United Kingdom (94.8%), Germany (94.8%), and Finland (94.3%) [3]. Studies of people who do not use the internet on a voluntary basis, even though they live in a country with well-developed ICT, may be useful policy material for not only those countries but also ones that currently lack ICT.

When discussing the social problem of internet use, researchers have mainly talked about vulnerability related to various factors on the premise that the internet should be available to anyone if they want it. Nevertheless, if there are people who do not want to use the internet despite having the circumstances and conditions for using it, another policy consideration will be needed. According to the Pew Research Center's report on the five factors related to internet nonuse, 15% of Americans do not use the internet and 92% of adults who do not use the internet said that they were not interested in using the internet or email in the future [4]. In the survey, the following factors were related to digital divide: age, income and educational attainment, community type, disability, and Spanish-speaking preference. Previous studies that analyzed factors related to internet use have reported age, disability, income, and education level as related factors [5-8].

However, it can be said that the factors that impede the use of the internet and the factors related to the nonuse of the internet due to voluntary intention have different meanings. As shown in the findings above, it is likely that people who are not willing to use the internet, especially those who have various conditions to use the internet, may have different contexts in the matter of digital divide.

This attitude toward internet use may be a sign of a more fundamental social vulnerability and not just about the internet. For those who have uncommon behaviors for voluntary reasons like this, it is necessary to examine the social and psychological situations they face.

High subjective well-being is associated with many desirable outcomes [9], such as positive development among young adults [10], healthier and longer lives [10], and democratic attitudes [11]. In a longitudinal study [12], the authors suggested that life satisfaction is prospectively associated with the occurrence of several major events in work and family life. Because life satisfaction is a trait-like construct [12,13], persons with higher satisfaction are more likable and more socially active [14] and are therefore more likely to have satisfying social relationships [12]. Additionally, there are research findings suggesting that life satisfaction motivates a greater willingness to engage in preventive health care, and these behaviors have been reported [15-18]. As such, life satisfaction not only has meaning as a result of certain life events, but also is a factor influencing an individual's attitude and behavior.

Social capital is a source of empowerment [19,20] and interacts with policy implementation [21]. Social capital can affect individuals' risk for negative health outcomes and their engagement in risk behaviors [22]. Putnam [23] has explained the distinction between bridging and bonding social capitals. Bonding social capital is inward looking and tends to reinforce exclusive identities and homogeneous groups. On the other hand, bridging social capital is outward looking and

encompasses people across diverse social cleavages. A study that was conducted to explore the relationship between social capital and willingness to join a community-based health insurance program found that social capital was a key determinant of willingness to join the program, and that people with a high level of bonding capital would be reluctant to cooperate with others [24]. In a study that explored the literature on social capital, human capital, and social learning theory, the author raised the hypothesis that the decline in social capital is negatively influencing recent college graduates' formation of soft skills. It was also suggested that college students gain the cultural and behavioral information and sensitivity they need to learn soft skills through the process of building social capital [25]. Additionally, social capital has a significant effect on information-related behaviors, particularly on the choice of source [26].

Thus, it can be assumed that regardless of the abundance of material resources, a poor level of subjective well-being or social capital can lead to a vicious cycle that inhibits universal and positive social behaviors. This study aimed to find out how life satisfaction and social capital act on voluntary nonusers of the internet, who live in South Korea.

Methods

Study Participants

We used data from the 2018 Survey on the Actual Condition of the Digital Divide conducted in South Korea. Since 2002, this survey has been annually conducted by the Ministry of Science and ICT along with the National Information Society Agency (NIA). It is done in order to check the annual performance of the digital information disparity policy through time-series research and analysis, as well as to provide the basic data necessary for future effective policy implementation. The survey consists of the following two parts: (1) digital informatization level and (2) attitudes toward information use. The data for each year of the survey were collected via a cross-sectional and nationally representative survey using a multistage stratified sampling design. After obtaining informed consent from participants, the health interview was performed by trained interviewers of the research corporation Kantar Korea. The survey was approved by the institutional review board at the NIA.

This survey can also be found in another research paper [27]. The survey data for 2018 included information for 7000 individuals. After the exclusion of those who were aged ≤18 years, 6150 participants remained in the study.

Assessment

Age, gender, education, occupation, household income, family size, and community type were included as sociodemographic variables (Table 1). The variables were categorized as follows: age (19-29, 30-39, 40-49, 50-59, or ≥60 years); education level (elementary school or lower, middle school, high school, or college graduate or higher); occupation (office/sales/technical/professional, agriculture/forestry/fishery, simple labor, housewife, student, or not employed); average monthly household income (≤0.99, 1-1.99, 2-2.99, 3-3.99,

4-4.99, 5-5.99, or ≥6.0 million Korean won [KRW]/month) (1 KRW = 0.00089 USD); family size (single-person household or multi-person household); and community type (urban or rural). In the questionnaire for disability, whether the respondent was a registered disabled person was investigated.

To investigate the respondents' internet usage, the question, "When was the last time you recently used the internet?" was asked. The factors involved with this question included the use of PCs, mobile phones, tablet PCs, Internet Protocol televisions, etc, regardless of location. In response, respondents were asked to choose one of the following: "Within the last month," "I have not used the internet for more than a month," and "I have never used the internet." The respondents were asked to choose one of the following reasons why they did not use the internet: "I am willing to use it, but I am not able to because I am not able to use it (expensive rates, no digital devices to use, physical disabilities, difficulties in how to use it, etc)" or "I can use it, but I do not feel the need to use it, so I do not use it." For those who responded "I can use it, but I do not feel the need to use it, so I do not use it," the following additional question was asked: "What is the main reason you do not use the internet?." In response, respondents were asked to choose one of the following: "Because there is no inconvenience without using the internet," "I have a negative thought about using the internet (wasting time, not very helpful, etc)," "Because of the reluctance to learn and use new things," "Because most people who are close to me do not use the internet," and "I am concerned about spam mail, personal information leakage, illegal information distribution, internet addiction, etc."

Social capital was measured using the Internet Social Capital Scale (ISCS) [28], which was developed based on the conceptualization of bonding and bridging social capital by Putnam [23]. The ISCS consists of 10 items for bonding social capital and 10 items for bridging social capital. A 4-point Likert scale instead of the original 5-point scale was used for each of the measurement items, ranging from 1 (strongly disagree) to 4 (strongly agree). The Cronbach alpha coefficient for this scale was .898.

In order to investigate satisfaction in various aspects of daily life, satisfaction was examined in the following eight areas: leisure and cultural life, economic conditions, social activities (community, gathering, community participation, etc), interpersonal relationships, family relationships, what I do (such as academic or work activities), health status, and politics or government policy. The daily life satisfaction was measured by having respondents choose "not satisfied at all," "unsatisfied," "satisfied," or "very satisfied" as an answer to the question, "How satisfied are you with the items below in your daily life?." The Cronbach alpha coefficient for this scale was .797.

Life satisfaction was assessed using the Satisfaction with Life Scale (SWLS), which was developed by Diener et al [29]. The SWLS consists of five items that measure satisfaction with one's life. Participants were asked to rate each item on a 5-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), and the total score could range from 5 to 35. The total scores of the scale can be interpreted as "extremely dissatisfied" (5-9), "dissatisfied" (10-14), "slightly below average in life satisfaction" (15-19), "average score" (20-24), "high score" (25-29), or "very high score (highly satisfied with life)" (30-35). The Korean version was obtained from the public database of the original author [30]. The Cronbach alpha coefficient for this scale was .887.

Table 1. Characteristics of the sample (N=6150).

Characteristic	Value, n (%) or mean (SD)
Gender (missing: 99, 1.61%)	
Male	3075 (50.82)
Female	2976 (49.18)
Age (years)	
19-29	1033 (16.80)
30-39	864 (14.05)
40-49	1015 (16.50)
50-59	1531 (24.89)
≥60	1707 (27.76)
Education	
Elementary school or lower	452 (7.35)
Middle school	725 (11.79)
High school	3000 (48.78)
College or higher	1973 (32.08)
Household income (million KRW[a]/month)	
≤0.99	226 (3.67)
1-1.99	706 (11.48)
2-2.99	1025 (16.67)
3-3.99	1496 (24.33)
4-5.99	2254 (36.65)
≥6	432 (7.02)
Occupation	
Office, sales, technical, or professional	3935 (63.98)
Agriculture, forestry, or fishery	175 (2.85)
Simple labor	189 (3.07)
Student	390 (6.34)
Housewife	1107 (18.00)
Not employed	354 (5.76)
Family size	
Single-person household	516 (8.39)
Multi-person household	5634 (91.61)
Community type	
Urban	4642 (75.48)
Rural	1508 (24.52)
Disability (yes)	70 (1.14)
Internet usage	
Voluntary nonuser	346 (5.63)
Others	5804 (94.37)
Bonding social capital score, mean (SD)	27.09 (4.61)
Bridging social capital score, mean (SD)	27.85 (5.44)
Satisfaction in various aspects of daily life	
Leisure and cultural life (satisfied)	4442 (72.23)

Characteristic	Value, n (%) or mean (SD)
Economic conditions (satisfied)	3592 (58.41)
Social activities (community, gathering, community participation, etc; satisfied)	4099 (66.65)
Interpersonal relationships (satisfied)	5149 (83.72)
Family relationships (satisfied)	5516 (89.69)
What I do (such as academic and work activities; satisfied)	4439 (72.18)
Health status (satisfied)	4602 (74.83)
Politics or government policy (satisfied)	2999 (48.76)
Life satisfaction	
Equal or higher than average	4079 (66.33)
Lower than average	2071 (33.67)

[a]KRW: Korean won.

Statistical Analysis

In this study, a voluntary nonuser of the internet was defined as a person who answered in the questionnaire that he or she had never used the internet and responded that he or she had the condition or ability to use the internet but did not feel the need to use it.

The daily life satisfaction categories of "not satisfied at all" and "unsatisfied" were grouped as "not satisfied," and the categories of "satisfied" and "very satisfied" were grouped as "satisfied." Accordingly, they were used as binary variables for the chi-square test and multiple logistic regression.

The SWLS categories of "extremely dissatisfied" (5-9), "dissatisfied" (10-14), and "slightly below average in life satisfaction" (15-19) were categorized as "lower than average," and the categories of "average score" (20-24), "high score" (25-29), and "very high score or highly satisfied with life" (30-35) were categorized as "equal or higher than average." Accordingly, they were used as binary variables for the chi-square test and multiple logistic regression.

To explore crude associations between potentially explanatory variables, such as sociodemographic variables, disability, daily life satisfaction, life satisfaction, and the act of voluntarily not using the internet, the Pearson chi-square test was conducted. A Student t test was used to compare the mean of bonding social capital and bridging social capital between the two groups based on whether individuals were not voluntarily using the internet. Multiple logistic regression explored the predicting factors of the act of voluntarily not using the internet. Variables with a significance level of $P<.05$ in the bivariate analysis were entered into the multivariate model. The analysis was performed using IBM SPSS Version 16 software (IBM Corp).

Ethics Approval

This study was approved by the NIA Institutional Review Board. All of the participants provided written informed consent for the survey. The funding source had no role in writing or submitting this research paper.

Results

Table 1 shows the general characteristics of the participants. A total of 346 respondents (N=6150, 5.63%) represented voluntary nonusers of the internet. In most cases, the education level was above middle school. Overall, 69.90% (4299/6150) of respondents were employed and 18.00% (1107/6150) were housewives. Moreover, 24.52% (1508/6150) of the respondents were living in rural areas and 1.14% (70/6150) were registered as being disabled.

The mean score of bonding social capital was similar to that of bridging social capital. In terms of satisfaction in various aspects of daily life, the fraction of those satisfied was the highest for family relationships (5516/6150, 89.69%) and the lowest for politics or government policies (2999/6150, 48.76%). Regarding the SWLS score, 66.33% (4079/6150) of respondents had an "equal or higher than average" score.

For the question, "What is the main reason you do not use the internet?," the frequencies of responses were as follows: "Because there is no inconvenience without using the internet," 82.1%; "I have a negative thought about using the internet (wasting time, not very helpful, etc)," 5.5%; "Because of the reluctance to learn and use new things," 7.8%; "Because most people who are close to me do not use the internet," 3.2%; "I am concerned about spam mail, personal information leakage, illegal information distribution, internet addiction, etc," 1.2%; and "other," 0.2%.

When the association between various factors and internet usage was assessed, the variables that showed statistical significance were gender, age, education, household income, occupation, family size, community type, disability, bonding social capital, bridging social capital, daily life satisfaction (in every aspect), and life satisfaction (SWLS) (Table 2).

Although there are limitations in interpretation because the approach is a univariate analysis, the percentage of people who did not voluntarily use the internet was higher among women, elderly people, less educated people, people with low income, unemployed people, agriculture/forestry/fishery workers, simple laborers, housewives, people in single-person households, and disabled people.

For people who did not voluntarily use the internet, the bonding social capital and bridging social capital scores were lower and the score difference was greater for bridging social capital. Among those who were not satisfied in various aspects of life, the percentage of people who did not voluntarily use the internet was higher. This trend was the same in the case of life satisfaction measured by the SWLS.

A multivariate logistic regression analysis was conducted to explore the predictors of the behavior of not voluntarily using the internet (Table 3). We reviewed the odds ratio (OR) and CI of each calculation. Age, education level, bonding and bridging social capitals, and daily life satisfaction for health status were found to be factors related to the behavior of not voluntarily using the internet. However, gender, household income, occupation, family size, and community type were not related to voluntary internet nonuse.

As age increased by 1 year, the likelihood of not voluntarily using the internet increased 1.14 times (OR 1.14, 95% CI 1.12-1.17). A higher level of education was associated with less likelihood that people would not voluntarily use the internet. The likelihood that a person with a higher education level than a university degree would not voluntarily use the internet was 0.14 times the likelihood for a person with a lower education level than middle school (OR 0.14, 95% CI 0.05-0.45). In the case of social capital, bonding social capital and bridging social capital showed different directions for internet use behavior. As the score of bonding social capital increased by 1, the likelihood of not voluntarily using the internet increased by 1.04 times (OR 1.04, 95% CI 1.01-1.08). On the other hand, the likelihood of not using the internet voluntarily decreased by 0.93 times as the score of bridging social capital increased by 1 (OR 0.93, 95% CI 0.91-0.96). Among the many aspects of everyday life, those who were not satisfied with their health status were 1.38 times more likely to not voluntarily use the internet than those who were satisfied (OR 1.38, 95% CI 1.01-1.87).

Table 2. Association between various factors and life satisfaction.

Characteristic	Internet usage, n (%) or mean (SD)		P value
	Voluntary nonuser	Others	
Gender			<.001[a]
Male (n=3075)	114 (3.7)	2961 (96.3)	
Female (n=2976)	223 (7.5)	2753 (92.5)	
Age (years)			<.001[a]
19-29 (n=1033)	0 (0.0)	1033 (100.0)	
30-39 (n=864)	0 (0.0)	864 (100.0)	
40-49 (n=1015)	0 (0.0)	1015 (100.0)	
50-59 (n=1531)	12 (0.8)	1519 (99.2)	
≥60 (n=1707)	334 (19.6)	1373 (80.4)	
Education			<.001[a]
Elementary school or lower (n=452)	160 (35.4)	292 (64.6)	
Middle school (n=725)	121 (16.7)	604 (83.3)	
High school (n=3000)	61 (2.0)	2939 (98.0)	
College or higher (n=1973)	4 (0.2)	1969 (99.8)	
Household income (million KRW[b]/month)			<.001[a]
≤0.99 (n=226)	85 (37.6)	141 (62.4)	
1-1.99 (n=706)	141 (20.0)	565 (80.0)	
2-2.99 (n=1025)	61 (6.0)	964 (94.0)	
3-3.99 (n=1496)	33 (2.2)	1463 (97.8)	
4-5.99 (n=2254)	22 (1.0)	2232 (99.0)	
≥6 (n=432)	4 (0.9)	428 (99.1)	
Occupation			<.001[a]
Office, sales, technical, or professional (n=3935)	78 (2.0)	3857 (98.0)	
Agriculture, forestry, fishery (n=175)	24 (13.7)	151 (86.3)	
Simple labor (n=189)	19 (10.1)	170 (89.9)	
Student (n=390)	0 (0.0)	390 (100.0)	
Housewife (n=1107)	139 (12.6)	968 (87.4)	
Not employed (n=354)	86 (24.3)	268 (75.7)	
Family size			<.001[a]
Single-person household (n=516)	88 (17.1)	428 (82.9)	
Multi-person household (n=5634)	258 (4.6)	5376 (95.4)	
Community type			.01[a]
Urban (n=4642)	241 (5.2)	4401 (94.8)	
Rural (n=1508)	105 (7.0)	1403 (93.0)	
Disability			<.001[a]
Yes (n=70)	12 (17.1)	58 (82.9)	
No (n=6080)	334 (5.5)	5746 (94.5)	
Bonding social capital score, mean (SD)	25.07 (4.81)	27.21 (4.57)	<.001[c]
Bridging social capital score, mean (SD)	23.25 (6.17)	28.12 (5.27)	<.001[c]

Characteristic	Internet usage, n (%) or mean (SD)		*P* value
	Voluntary nonuser	Others	
Satisfaction in various aspects of daily life			
Leisure and cultural life			<.001[a]
Satisfied (n=4442)	149 (3.4)	4293 (96.6)	
Dissatisfied (n=1708)	197 (11.5)	1511 (88.5)	
Economic conditions			<.001[a]
Satisfied (n=3592)	111 (3.1)	3481 (96.9)	
Dissatisfied (n=2558)	235 (9.2)	2323 (90.8)	
Social activities (community, gathering, community participation, etc)			<.001[a]
Satisfied (n=4099)	163 (4.0)	3936 (96.0)	
Dissatisfied (n=2051)	183 (8.9)	1868 (91.1)	
Interpersonal relationships			<.001[a]
Satisfied (n=5149)	244 (4.7)	4905 (95.3)	
Dissatisfied (n=1001)	102 (10.2)	899 (89.8)	
Family relationships			<.001[a]
Satisfied (n=5516)	289 (5.2)	5227 (94.8)	
Dissatisfied (n=634)	57 (9.0)	577 (91.0)	
What I do (such as academic and work activities)			<.001[a]
Satisfied (n=4439)	172 (3.9)	4267 (96.1)	
Dissatisfied (n=1711)	174 (10.2)	1537 (89.8)	
Health status			<.001[a]
Satisfied (n=4602)	139 (3.0)	4463 (97.0)	
Dissatisfied (n=1548)	207 (13.4)	1341 (86.6)	
Politics or government policy			<.001[a]
Satisfied (n=2999)	104 (3.5)	2895 (96.5)	
Dissatisfied (n=3151)	242 (7.7)	2909 (92.3)	
Life satisfaction			<.001[a]
Equal or higher than average (n=4079)	159 (3.9)	3920 (96.1)	
Lower than average (n=2071)	187 (9.0)	1884 (91.0)	

[a]*P* value determined using the Pearson chi-square test.

[b]KRW: Korean won.

[c]*P* value determined using Student *t* test.

Table 3. Multiple logistic regression results examining the factors associated with the act of voluntarily not using the internet.

Characteristic	Voluntary nonusers	
	Adjusted OR[a]	95% CI
Gender (female)	1.14	0.78-1.66
Age (continuous)	1.14	1.12-1.17
Education		
Elementary school or lower	Reference	N/A[b]
Middle school	0.82	0.59-1.15
High school	0.43	0.28-0.66
College or higher	0.14	0.05-0.45
Household income (million KRW[c]/month)		
≤0.99	Reference	N/A
1-1.99	0.85	0.55-1.33
2-2.99	0.73	0.43-1.24
3-3.99	0.73	0.39-1.36
4-5.99	0.58	0.29-1.15
≥6	0.64	0.20-2.08
Occupation		
Office, sales, technical, or professional	Reference	N/A
Agriculture, forestry, fishery	1.06	0.57-1.97
Simple labor	1.01	0.56-1.83
Student	__d,e	__d,e
Housewife	1.20	0.80-1.82
Not employed	1.25	0.81-1.94
Family size		
Single-person household	Reference	N/A
Multi-person household	0.88	0.58-1.33
Community type		
Urban	Reference	N/A
Rural	0.91	0.66-1.27
Disability		
No	Reference	N/A
Yes	0.76	0.35-1.66
Bonding social capital score (continuous)	1.04	1.01-1.08
Bridging social capital score (continuous)	0.93	0.91-0.96
Satisfaction in various aspects of daily life		
Leisure and cultural life		
Satisfied	Reference	N/A
Dissatisfied	1.22	0.88-1.68
Economic conditions		
Satisfied	Reference	N/A
Dissatisfied	1.16	0.82-1.64
Social activities (community, gathering, community participation, etc)		

Characteristic	Voluntary nonusers	
	Adjusted OR[a]	95% CI
Satisfied	Reference	N/A
Dissatisfied	0.85	0.62-1.17
Interpersonal relationships		
Satisfied	Reference	N/A
Dissatisfied	1.10	0.77-1.56
Family relationships		
Satisfied	Reference	N/A
Dissatisfied	1.00	0.68-1.47
What I do (such as academic and work activities)		
Satisfied	Reference	N/A
Dissatisfied	0.85	0.62-1.16
Health status		
Satisfied	Reference	N/A
Dissatisfied	1.38	1.01-1.87
Politics or government policy		
Satisfied	Reference	N/A
Dissatisfied	0.96	0.69-1.33
Life satisfaction		
Equal or higher than average	Reference	N/A
Lower than average	1.20	0.88-1.63

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]KRW: Korean won.

[d]For students, the number of voluntary internet nonusers is zero.

[e]Not available.

Discussion

The purpose of this study was to explore the factors related to the behavior of people who do not voluntarily use the internet, regardless of conditions. As factors related to such voluntary behavior regardless of conditions, age and education level, which were previously reported to be related to internet use, have been reaffirmed. However, household income and disability were not related to the act of voluntarily not using the internet regardless of the circumstances. These results suggest that spontaneity related to the nonuse of the internet is not related to income or disability; yet, spontaneity is affected when age increases or the level of education is low. These results mean that countries or communities that promote an informatized society should help the elderly or less educated to easily understand the benefits of the internet that they have never experienced and increase their ability to use it. Interestingly, it can be seen that social capital exerts different influences on the voluntary nonuse of the internet, depending on its attributes. The finding that a higher bonding social capital is associated with higher voluntary nonuse of the internet suggests that the unmet needs are somewhat being met by the support or help of

others, even if the individuals do not use the internet in their own group. On the other hand, the finding that a higher bridging social capital is associated with lower voluntary internet nonuse can be seen as a natural result reflecting the interest and willingness in communicating with other groups outside of their own group. As a strategy for voluntary internet nonusers who are reluctant to use the internet, it can be considered to promote campaigns targeting existing internet users in order to ensure that changes in attitudes and behaviors of people with low bridging social capital can occur owing to the influence of people with high levels of ICT use in the groups.

In terms of life satisfaction, only satisfaction with health status was related to voluntary internet nonuse. It may not be something to be taken lightly that a person who is not satisfied with his or her health does not want to use the internet. Those who are not satisfied with their health owing to poor objective or subjective health conditions may find various communication or participation activities, and internet use annoying. For these people, quality of life can worsen by entering a vicious circle.

In this study, it was found that sociodemographic factors, such as age and education level, which are difficult to modify, and psychosocial factors located deeper than the visible living

XSL•FO
RenderX

conditions, such as social capital and life satisfaction, are involved in voluntary internet nonuse. These results also suggest that it is not desirable to proceed with ICT-related policies on a separate track, but rather that they should be comprehensively approached together with other social policies that design various social interventions in order to enhance equity within the society. The results of this study suggest that in order to fully extract the roots of the weed called digital divide, it is necessary to consider the sociopsychological soil that can sustain the roots. The recent COVID-19 pandemic of 2019 has led many countries to consider the activation of telemedicine or telehealth [31,32]. Compared with the rapidly changing global health care situation and the development of ICT, it is time to carefully review whether the social welfare infrastructure of each country can keep pace.

This study has several strengths. First, it used nationally representative data gathered by a professional research firm using random probability allocation. Second, relatively diverse variables related to internet usage were investigated and used for multivariate analysis. Third, this study was conducted in a country with well-equipped ICT infrastructure. This research environment is suitable for investigating relevant factors for voluntary internet users regardless of conditions, and it can provide useful information to many countries that have built or are building a similar level of ICT infrastructure. On the other hand, this study has some limitations. This study is a cross-sectional study (not a longitudinal one), and so, there are limitations regarding the interpretation of the study results. Moreover, no detailed psychological investigations were carried out to determine which psychological mechanisms exist between low satisfaction with health status and voluntary internet nonuse. In the future, it seems necessary to produce more thorough evidence utilizing community intervention studies targeting the vulnerable groups of the internet, including voluntary internet nonusers.

Conflicts of Interest

None declared.

References

1. Perron BE, Taylor HO, Glass J, Margerum-Leys J. Information and Communication Technologies in Social Work. ASW 2010 Mar 18;11(1):67-81. [doi: 10.18060/241]
2. Gross domestic product 2019. World Bank. URL: https://databank.worldbank.org/data/download/GDP.pdf [accessed 2020-10-29]
3. ICT Access and Usage by Households and Individuals. OECD iLibrary. URL: https://www.oecd-ilibrary.org/science-and-technology/data/oecd-telecommunications-and-internet-statistics/ict-access-and-usage-by-households-and-individuals_b9823565-en [accessed 2020-11-05]
4. Who's not online? 5 factors tied to the digital divide. Pew Research Center. 2013. URL: https://www.pewresearch.org/fact-tank/2013/11/08/whos-not-online-5-factors-tied-to-the-digital-divide/ [accessed 2020-11-05]
5. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
6. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
7. Gounopoulos E, Kazanidis I, Valsamidis S, Kontogiannis S. A Hierarchical Analysis of Internet Adoption and Usage by the Seniors. In: Janowicz-Lomott M, Łyskawa K, Polychronidou P, Karasavvoglou A, editors. Economic and Financial Challenges for Balkan and Eastern European Countries. Cham: Springer; Apr 28, 2020.
8. Cruz-Cunha MM, Miranda IM, Gonçalves P. Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care. Hershey, PA: IGI Global; 2013.
9. Calvo R, Zheng Y, Kumar S, Olgiati A, Berkman L. Well-being and social capital on planet earth: cross-national evidence from 142 countries. PLoS One 2012;7(8):e42793 [FREE Full text] [doi: 10.1371/journal.pone.0042793] [Medline: 22916160]
10. Park N. The Role of Subjective Well-Being in Positive Youth Development. The ANNALS of the American Academy of Political and Social Science 2016 Sep 08;591(1):25-39. [doi: 10.1177/0002716203260078]
11. Tov W, Diener E. The well-being of nations: Linking together trust, cooperation, and democracy. In: Sullivan BA, Snyder M, Sullivan JA, editors. Cooperation: The political psychology of effective human interaction. Malden, MA: Blackwell; 2008:155-173.
12. Luhmann M, Lucas RE, Eid M, Diener E. The Prospective Effect of Life Satisfaction on Life Events. Social Psychological and Personality Science 2012 Mar 20;4(1):39-45. [doi: 10.1177/1948550612440105]
13. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: Three decades of progress. Psychological Bulletin 1999;125(2):276-302. [doi: 10.1037/0033-2909.125.2.276]
14. Okun MA, Stock WA, Haring MJ, Witter RA. The Social Activity/Subjective Well-Being Relation. Res Aging 2016 Aug 19;6(1):45-65. [doi: 10.1177/0164027584006001003]

XSL•FO

RenderX

15.   Kim ES, Kubzansky LD, Smith J. Life satisfaction and use of preventive health care services. Health Psychol 2015 Jul;34(7):779-782 [FREE Full text] [doi: 10.1037/hea0000174] [Medline: 25420064]

16.   Grant N, Wardle J, Steptoe A. The relationship between life satisfaction and health behavior: a cross-cultural analysis of young adults. Int J Behav Med 2009;16(3):259-268. [doi: 10.1007/s12529-009-9032-x] [Medline: 19319695]

17.   Wickrama KA, Ralston PA, O'Neal CW, Ilich JZ, Harris CM, Coccia C, et al. Linking Life Dissatisfaction to Health Behaviors of Older African Americans Through Psychological Competency and Vulnerability. Res Aging 2012 Jun 13;35(5):591-611. [doi: 10.1177/0164027512449473]

18.   Grant N, Wardle J, Steptoe A. The relationship between life satisfaction and health behavior: a cross-cultural analysis of young adults. Int J Behav Med 2009 Mar 25;16(3):259-268. [doi: 10.1007/s12529-009-9032-x] [Medline: 19319695]

19.   Ansari S, Munir K, Gregg T. Impact at the 'Bottom of the Pyramid': The Role of Social Capital in Capability Development and Community Empowerment. Journal of Management Studies 2012 Jan 10;49(4):813-842. [doi: 10.1111/j.1467-6486.2012.01042.x]

20.   Janssens W. Women's Empowerment and the Creation of Social Capital in Indian Villages. World Development 2010 Jul;38(7):974-988 [FREE Full text] [doi: 10.1016/j.worlddev.2009.12.004]

21.   Compton ME. Less Bang for Your Buck? How Social Capital Constrains the Effectiveness of Social Welfare Spending. State Polit Policy Q 2018 Sep;18(3):215-245 [FREE Full text] [doi: 10.1177/1532440018775424] [Medline: 30595673]

22.   Kaljee LM, Chen X. Social capital and risk and protective behaviors: a global health perspective. Adolesc Health Med Ther 2011 Dec 17;2011(2):113-122 [FREE Full text] [doi: 10.2147/AHMT.S26560] [Medline: 23243387]

23.   Putnam R. Bowling Alone: The Collapse and Revival of American Community. New York, NY: Simon and Schuster; 2000.

24.   Ko H, Kim H, Yoon CG, Kim CY. Social capital as a key determinant of willingness to join community-based health insurance: a household survey in Nepal. Public Health 2018 Jul;160:52-61 [FREE Full text] [doi: 10.1016/j.puhe.2018.03.033] [Medline: 29734013]

25.   Andreas S. Effects of the decline in social capital on college graduates' soft skills. Industry and Higher Education 2018 Jan 03;32(1):47-56. [doi: 10.1177/0950422217749277]

26.   Johnson CA. Social capital and the search for information: Examining the role of social capital in information seeking behavior in Mongolia. J. Am. Soc. Inf. Sci 2007 Apr;58(6):883-894. [doi: 10.1002/asi.20561]

27.   Park S, Jae Kim G. Lessons from South Korea's digital divide index (DDI). INFO 2014 May 06;16(3):72-84. [doi: 10.1108/info-07-2013-0042]

28.   Williams D. On and Off the 'Net: Scales for Social Capital in an Online Era. J Comp Mediated Comm 2006 Jan;11(2):593-628. [doi: 10.1111/j.1083-6101.2006.00029.x]

29.   Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]

30.   Satisfaction with Life Scale (SWLS). Ed Diener. URL: https://eddiener.com/scales/7 [accessed 2020-10-29]

31.   Bloem BR, Dorsey ER, Okun MS. The Coronavirus Disease 2019 Crisis as Catalyst for Telemedicine for Chronic Neurological Disorders. JAMA Neurol 2020 Aug 01;77(8):927-928. [doi: 10.1001/jamaneurol.2020.1452] [Medline: 32329796]

32.   Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun 20;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]

Abbreviations

ICT: information and communications technology
ISCS: Internet Social Capital Scale
NIA: National Information Society Agency
OR: odds ratio
SWLS: Satisfaction with Life Scale

Original Paper

Internet Use in Old Age: Results of a German Population-Representative Survey

Janine Quittschalle[1], MA; Janine Stein[1], PhD; Melanie Luppa[1], PhD; Alexander Pabst[1], PhD; Margrit Löbner[1], PhD; Hans-Helmut Koenig[2], MPH, MD, Prof Dr; Steffi G Riedel-Heller[1], MPH, MD, Prof Dr

[1]Institute of Social Medicine, Occupational Health and Public Health (ISAP), University of Germany, Leipzig, Germany

[2]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Janine Quittschalle, MA
Institute of Social Medicine, Occupational Health and Public Health (ISAP)
University of Germany
Philipp-Rosenthal-Straße 55, Building W
Leipzig, 04103
Germany
Phone: 49 341 97 24571
Email: Janine.Quittschalle@medizin.uni-leipzig.de

Abstract

Background: The internet has the potential to foster healthy lifestyles and to support chronic disease management. Older adults could benefit from using the internet and other information and communication technology to access health-related information and interventions available online.

Objective: The aim of this study was to investigate factors influencing internet use in older and oldest age groups and to determine the frequency of internet use for health-related purposes.

Methods: Using data from a nationally representative telephone survey of older adults aged 75 years and over, a sample of 999 people was assessed using structured clinical interviews. Descriptive and binary logistic regression analyses were performed.

Results: Overall, 42.6% (418/999) of participants used the internet. Among those, 55.7% (233/417) searched the internet for health-related information. Regression analyses revealed that internet use was significantly associated with younger age (odds ratio [OR] 0.89, 95% CI 0.85-0.92; $P<.001$), male gender (OR 2.84, 95% CI 2.02-4.00; $P<.001$), higher education levels (OR 6.69, 95% CI 4.48-9.99; $P<.001$), a wider social network (OR 1.04, 95% CI 1.01-1.07; $P=.01$), higher health-related quality of life (OR 1.02, 95% CI 1.00-1.03; $P=.006$), lower levels of depressive symptoms (OR 0.89, 95% CI 0.80-0.99; $P=.04$), and higher rates of chronic illness (OR 1.12, 95% CI 1.04-1.21; $P<.004$).

Conclusions: This study provides population-representative data on internet use in old age in Germany. People in the older and oldest age groups participate in online activities. Understanding the factors that are associated with older adults internet use can contribute to developing tailored interventions and eHealth (electronic health) services to improve well-being in older adults.

Introduction

Background

Aging is often associated with major changes and stressful life transitions (eg, shrinking social network, relocation, retirement) that have a big impact on well-being, health, and everyday life [1]. Researchers have highlighted the enormous benefits of internet and information and communication technology (ICT) use for the aging population. For instance, these benefits include enhancement of social connectivity [2], prevention or reduction of social isolation [3], increased information about offline leisure and recreational activities [4,5], and increased empowerment through access to information (eg, health-related issues) [6,7]. Various studies have explored factors at the individual level as well as factors of the social context that predict internet use in older age. Results have consistently shown that factors that are positively associated with internet use over 65 are younger age,

male gender, higher educational level, married or living with someone else, specifically partners or family members [2,8-12].

However, some researchers have found fault with these findings, arguing that many studies aggregated people of a certain age together into 1 homogenous category [13], or that they did not include older age groups (eg, 70+) [12]. Few studies that focus on older adults (75+) [4,8,12,13] have analyzed sociodemographic correlates, health determinants, and social context correlates of internet use. They did, however, confirm previous results, for example, with regard to age, gender, education, and marital status. Importantly, the authors found that adults aged 75 and older use the internet less frequently than younger seniors, which led them to conclude that the "digital divide" between this age group and other population groups is more pronounced [12,14]. Because of the benefits of digital technologies for older adults, especially for maintaining a healthy lifestyle [15,16], the digital divide by age is worrying. In order to enable future planning of the health care system, reliable information on internet usage behavior and its predictors in the older and oldest age groups is becoming increasingly important. In Germany, apart from continuous surveys on general internet use [17,18] and few studies on internet use for health-related purposes [19,20], there is a lack of research investigating the internet use of the older and oldest German population.

There is a yearly survey conducted by a national television broadcaster using telephone interviews to examine internet use and communication in Germany. The 2018 survey of online behavior was conducted with a sample of 2009 German-speaking persons aged 14 and older [17]. However, this study did not examine the determinants of internet use nor did it focus on older and oldest aged individuals. Therefore, the aim of this study is to investigate the frequency of internet use and factors associated with internet use among the German older and oldest aged 75-99. Based on a German population-representative sample, this study focuses on the following research questions:

- Do older and oldest German adults use the internet?
- Do they search the internet for physical or psychological health information?
- Are there differences between internet users and nonusers in old age?
- What are the correlates of internet use in older and oldest age?

Conceptual Framework

In order to structure the results of this study, we relied on Anderson's behavioral model of health service use [21] and applied this conceptual framework to internet usage. The behavioral model [21] contains 3 main underlying factors: (1) "Predisposing variables" which include demographic and relationship characteristics as well as health perceptions; (2) "Need variables" or the direct cause of service use including the amount and type of disease burden or strain (stressors); (3) "Enabling variables" which are both personal/social resources and community-based resources. This assumes that individuals have the resources and knowledge to know about and use appropriate services. These services must also be available.

Methods

Study Design and Sample

We analyzed data derived from the project "Needs assessment in the oldest-old: Application, psychometric examination and establishment of the adapted German version of the Camberwell Assessment of Need for the Elderly (CANE)," which was funded by the German Research Foundation (DFG). Within the project, a population-representative telephone survey was carried out on behalf of the University of Leipzig by a leading market, opinion, and social research institute (USUMA GmbH) in Germany between July and October 2016. The sampling method was based on a procedure implemented by the Association of German Market and Social Research Agency (AMD) and included registered and nonregistered telephone numbers of households throughout Germany. During the first step, individuals aged 75 years and older who lived in randomly selected households were identified as a contact person. In the case of more than 1 person aged 75 years and older living in the contacted household, the kish selection grid was applied in order to randomly select the target person within the household. This method is widely applied for selecting members within a household by using a preassigned table of random numbers to determine the target person. The landline telephone sample used to select the prospective households to be interviewed was converted into a representative population sample by weighting it according to key social demographic characteristics (age, gender, regional distribution by federal state). By applying this technique, equal probability of participation for each member of the randomly selected household, and therefore the representativeness of the sample, was ensured [22]. The detailed sample selection process is presented in Multimedia Appendix 1 and is reported in detail by Stein et al [22].

The survey conducted 1193/2823 (42.26%) telephone interviews. Interviews were excluded if dementia was suspected as assessed using the Six-item Cognitive Impairment Test (6-CIT) [23,24]. In the end, 1004 complete interviews with an average duration of 40 minutes were carried out. For the secondary analyses, the following inclusion criteria had to be met: (1) at least 75 years of age and (2) complete data on relevant variables were available. Additionally, 5 individuals were excluded from the study sample due to missing information or incomplete data on internet usage variables, resulting in an analytical sample size of 999 individuals.

Ethics Approval

The study was approved by the Ethics Committee of the Medical Faculty of the University of Leipzig. Prior to assessment, all study participants provided informed verbal consent.

Procedure and Instruments

The survey was conducted using standardized structured computer-assisted telephone interviews, which included sociodemographic variables and several standardized instruments. Prior to the survey, the instruments were tested and adapted within a pretest. Furthermore, the interviewers received comprehensive training.

Dependent Variables

The frequency of internet use was measured using a 5-point scale ranging from 1 to 5 (1=daily, 2=at least once a week, 3=at least once a month, 4=less than once a month, and 5=never). In addition, the frequency of internet searches for physical or psychological health information was assessed by using a 4-point scale ranging from 1 to 4 (1=yes, often; 2=yes, sometimes; 3=yes, but only rarely; and 4=no, never).

Independent Variables

Based on Anderson's behavioral model [21], explanatory variables were chosen. The predisposing characteristics include age, gender, marital status (married/with spouse, married/living apart, single, divorced, widowed), education, and living situation (living alone in a private household), living with others (spouse/partner, with other relatives, in a nursing home, assisted living situation, retirement home, other).

The need factors included depressive symptoms, health-related quality of life, number of chronic diseases, loneliness, and experiences of loss. Depressive symptoms were measured using the 15-item short German language version of the Geriatric Depression Scale (GDS) [25]. The GDS is used to identify symptoms of depression in older adults (eg, the basic satisfaction with one's life). The scale is a self-report instrument that uses a "Yes/No" format. The short version of the GDS showed good psychometric properties. Reliability coefficient estimated by Cronbach α was .81 for the GDS-15. [26]. According to Allen and Annells [27], a cut-off score on the GDS of 4/5 (noncase/case) corresponds to clinically relevant depressive symptoms and was therefore used in this study to screen participants for depression. The number of chronic diseases was measured by using a list of 22 chronic diseases (eg, high blood pressure, heart attack/coronary heart disease, lung disease, stroke, osteoporosis, diabetes, rheumatism, and cancer) with dichotomous response categories (yes/no). The participants were asked whether or not they had been recently diagnosed by a general practitioner with 1 or more of these chronic medical conditions. Health-related quality of life was assessed by using the Visual Analog Scale of the EuroQol 5-Dimensions 5-Level Questionnaire (EQ-5D-5L Questionnaire) or EQ-VAS [28]. The EQ-5D-5L Questionnaire assesses 5 dimensions of health states (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). With the vertical visual analogue scale (EQ-VAS) the participant is asked to gauge his or her current health state between best and worst imaginable. The EQ-VAS ranges from 0 (worst imaginable health state) to 100 (best imaginable health state). The EQ-VAS is frequently used as a quantitative measure of self-reported health state and has good psychometric properties in late life [29]. Psychometric properties of the German version of the EQ-5D-5L have been evaluated across different diseases [30-33]. Loneliness was assessed using the 3-item short version of the UCLA (University of California, Los Angeles) Loneliness Scale [34]. The scale captures subjective feelings of loneliness and social isolation (eg, the frequency of lack of company). Participants responded on a 3-point scale ranging from 1 (hardly ever) to 3 (often). The α coefficient of reliability is .72, indicating that the 3-item scale is a reliable measure for loneliness in a telephone survey [34].

Experiences of loss were measured using the Leipziger Lebensereignis-Liste (LLL). The LLL was adapted specifically for older adults and based on previously validated scales for the assessment of stressful life events: Social Readjustment Rating Scale [35], Recent Life Changes Questionnaire [36], and Life Events and Difficulties Schedule [37]. Participants were asked if they had experienced the "death of a significant other" within the last 12 months (yes/no), and if so, who died.

The enabling variables social network/social support were measured using the 6-item Lubben Social Network Scale (LSNS-6) [38]. The LSNS-6 consists of 6 questions assessing the size of the respondent's active social network (ie, number of relatives or friends seen or heard from ≥1 time per month), perceived social support (ie, number of relatives or friends who could be called for help), and perceived confidant network (ie, number of relatives or friends with whom the respondent could talk about private matters). Each LSNS-6 question is scored on a 0-5 scale. The total social network score is an equally weighted sum of these 6 questions, with scores ranging from 0 to 30. Higher scores indicate larger social networks or more frequent social contact. It has been demonstrated that the scale has good internal consistency (Cronbach α=.83) [38].

Statistical Analyses

Based on population statistics from the German Federal Statistics Office [39], data were weighted by USUMA in accordance with age, gender, and region. In order to obtain population-representative results, USUMA used design and adjustment weighting techniques [40]. In this study, unweighted absolute frequencies were presented, whereas any other analyses were performed and reported by using the weighting factor.

Statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS) version 24.0 for Windows (SPSS Inc.). In the analysis, the variable frequency of internet use was recoded into a dichotomous variable: *yes* and *no*. Individuals who reported never using the internet were classified as noninternet users. Individuals who reported using it less than once a month or more were classified as internet users. In addition, the variable frequency of internet searches for physical or psychological health information was recoded into a binary variable *yes* and *no*. Further, marital status was recoded into a binary variable (single vs married/with partner), and a new dichotomous variable *living situation* (living alone/living with others) was created. According to the international education classification "CASMIN" [41], the variable educational level was created with 3 categories (low, middle, and high). Finally, the variable "chronic diseases" was computed by counting the number of chronic conditions. This criterion was met if 75% or more of the 22 items on the list of chronic conditions were checked positive. If there were more missing values on the list (>15 unanswered items), the question was considered invalid (N=1).

Descriptive statistics presented are means and corresponding SDs, absolute frequencies, and percentages, as appropriate. The 2 groups (internet users/nonusers) were compared using independent samples *t* tests (unpaired) or chi-square tests as appropriate. Next, binary logistic regression analyses were conducted to examine factors associated with internet use in

older adults. The dichotomized variable internet use served as the dependent variable. The predictor variables included age, gender, education, marital status, domicile, experiences of social loss, GDS score, LSNS score, loneliness score, chronic disease, and EQ-5D-5L-VAS score. Furthermore, we performed a binary logistic regression analyses to investigate if health-related conditions are associated with internet search for physical or psychological health information within older internet users. The dichotomized variable internet search for physical or psychological health information (yes/no) served as the dependent variable. The predictor variables included GDS score and number of chronic disease. The significance level was set at $P \leq .05$ for all analyses.

Results

Participant Characteristics

Sociodemographic characteristics of the study sample and comparison between internet users and nonusers are presented in Table 1. The sample consisted of 999 individuals; 612 (59.1%) were female and 387 (40.9%) were male. The average age was 80.49, with a range of 75-99 years. More than one-third of the respondents had a low educational level (368/999, 36.3%), roughly one-third was highly educated (324/999, 31.3%), and 32.4% (307/999) had a middle educational level.

Table 1. Sociodemographic characteristics of the sample (N=999).

Variables	All (N=999, 100%)	Internet use		P-value
		Yes (n=418, 41.8%)	No (n=581, 58.2%)	
Age (years)				<.001
Mean (SD)	80.49 (4.69)	78.91 (3.57)	81.67 (5.07)	
Range	75-99	75-92	75-99	
Gender, n[a] (%[b])				<.001
Male	387 (40.9)	231 (57.5)	156 (26.6)	
Female	612 (59.1)	187 (42.5)	425 (73.4)	
Education[c], n (%)				<.001
Low	368 (36.3)	74 (17.7)	294 (50.1)	
Middle	324 (32.4)	142 (33.6)	182 (31.5)	
High	307 (31.3)	202 (48.7)	105 (18.4)	
Domicile[d], n (%)				<.001
Alone	572 (55.4)	185 (43.3)	387 (65.6)	
With spouse	378 (38.5)	218 (53.1)	160 (28.4)	
With relatives	46 (4.5)	17 (4.1)	29 (4.9)	
Other	14 (1.5)	4 (0.9)	10 (1.0)	
Missing	1 (0.1)	0 (0)	1 (0.1)	
Marital status, n (%)				<.001
Married, living together	353 (36.3)	203 (49.3)	150 (26.6)	
Married, living apart	25 (2.6)	15 (3.8)	10 (1.8)	
Single	74 (7.5)	26 (6.4)	48 (8.3)	
Divorced	77 (7.6)	28 (6.4)	49 (8.5)	
Widowed	469 (46.0)	145 (34.0)	324 (55.8)	
Missing	1 (0.1)	1 (0.1)	0 (0)	

[a]n: frequencies (none weight).

[b]%: percentages (weight).

[c]Educational classification according to the new CASMIN educational classification. Low: inadequately completed general education, general elementary education, basic vocational qualification or general elementary education, and vocational qualification; Middle: intermediate vocational qualification or intermediate general qualification and vocational qualification, intermediate general qualification, general maturity certificate, vocational maturity certificate/general maturity certificate, and vocational qualification; High: lower tertiary education—general diplomas/diplomas with vocational emphasis, higher tertiary education—lower level/higher level [41].

[d]Multiple responses possible.

Differences Between Internet Use and Nonuse

The majority of the sample consisted of internet nonusers (581/999, 57.4% vs 418/999, 42.6%, internet users). Of those using the internet, more than one-half searched the internet for physical or psychological health information (233/417, 55.7%). Figure 1 displays the frequency of internet use in general, as well as the frequency of internet use to obtain health-related information.

Figure 1. Frequency of internet use and frequency of internet use to obtain physical or psychological health information.
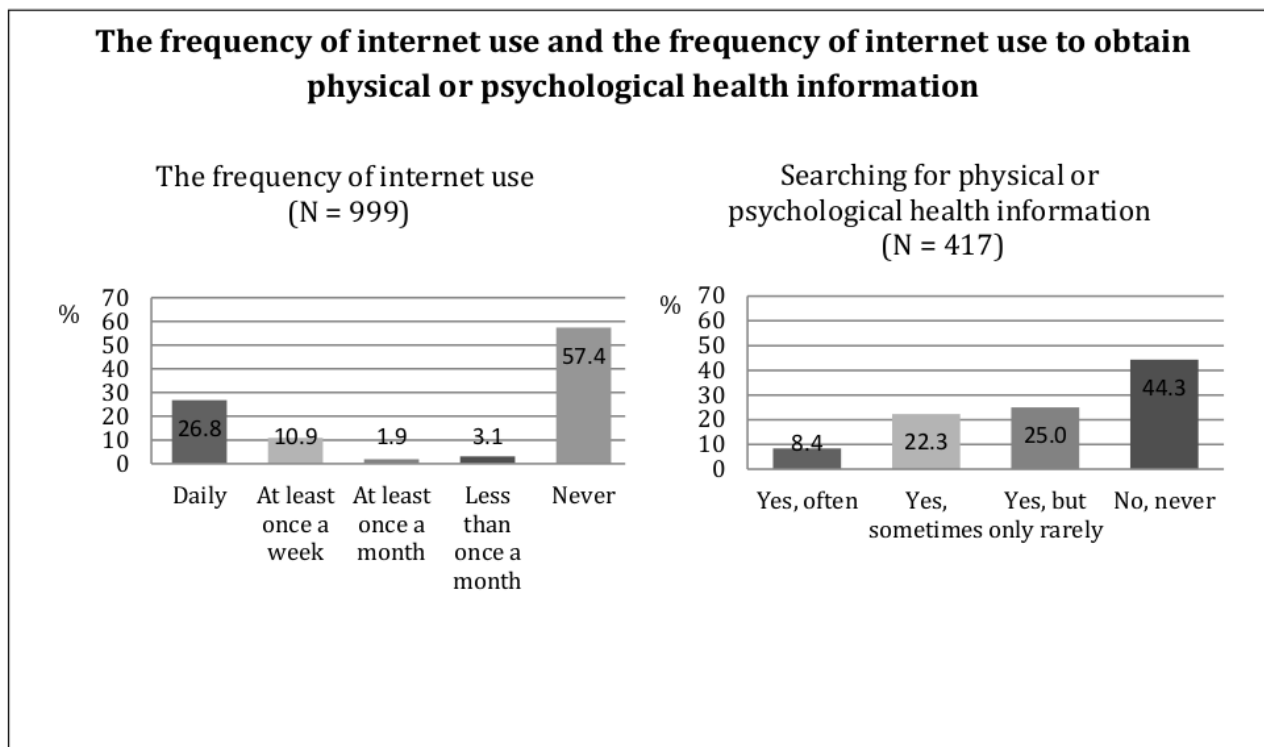
On a group level, there were several differences between users and nonusers, which are shown in Table 1 for sociodemographic variables and in Table 2 for social and health-related factors. In comparison to nonusers, older adults using the internet were more likely to be male ($P<.001$), tended to be younger ($P<.001$), were more likely to be middle or highly educated ($P<.001$), and were more likely to be married/with partner ($P<.001$) as well as to be living with their partner in the same household ($P<.001$).

With regard to social and health-related factors, individuals using the internet tended to have better self-rated health-related quality of life ($P<.001$), reported fewer depressive symptoms ($P<.001$), fewer chronic medical conditions ($P=.02$), and less loneliness ($P<.001$). Furthermore, internet users reported having a greater and more supportive social network than nonusers ($P<.001$). No significant differences were found regarding experiences of social loss ($P=.87$).

Table 2. Social and health outcomes of the sample (N=999).

Variables	All	Internet use		P-value
	(N=999, 100%)	Yes (n=418, 41.8%)	No (n=581, 58.2%)	
Health-related quality of Life VAS EQ-5D-5L[a]				<.001
Mean (SD)	73.46 (19.51)	78.02 (16.48)	70.08 (20.87)	
Range	0-100	10-100	0-100	
Missing, n[b] (%[c])	4 (0.4)	1 (0.3)	3 (0.4)	
UCLA Loneliness Scale[d]				<.001
Mean (SD)	3.57 (1.02)	3.41 (0.80)	3.68 (1.15)	
Range	3-9	3-9	3-9	
Missing, n (%)	15 (1.6)	4 (0.8)	11 (2.1)	
Experience of social loss				.87
Yes, n (%)	297 (30.0)	126 (30.2)	171 (29.8)	
No, n (%)	702 (70.0)	292 (69.8)	410 (70.2)	
GDS[e]				<.001
Mean (SD)	1.81 (2.07)	1.27 (1.58)	2.21 (2.30)	
Range	0-13	0-12	0-13	
Missing, n (%)	10 (1.0)	5 (1.0)	5 (0.9)	
LSNS[f]				<.001
Mean (SD)	15.90 (5.97)	17.57 (5.52)	14.64 (5.99)	
Range	0-30	0-30	1-30	
Missing, n (%)	39 (4.0)	12 (2.9)	27 (4.8)	
Chronic diseases[g]				.02
Mean (SD)	3.58 (2.33)	3.37 (2.13)	3.73 (2.46)	
Range	0-13	0-12	0-13	
Missing, n (%)	1 (0.1)	0 (0)	1 (0.2)	

[a]VAS EQ-5D-5L: Visual Analogue Scale of the EuroQol 5-Dimensions 5-Level Questionnaire; score ranges from 0 (worst imaginable health state) to 100 (best imaginable health state).

[b]n: frequencies (none weight).

[c]%: percentages (weight).

[d]UCLA: University of California, Los Angeles, Loneliness Scale; score ranges from 1 (hardly ever) to 3 (often).

[e]GDS: Geriatric Depression Scale; score ranges from 0 (no depression symptoms) to 15 (depression symptoms), cut-off: > 4.

[f]LSNS: Lubben Social Network Scale; score ranges from 0 to 30, with higher values reflecting more social networks and more social support.

[g]The sum score of chronic diseases ranges from 0 (no chronic conditions) to 22 (22 chronic conditions).

Predictors of Internet Use

A binary logistic regression was conducted to identify associations between several factors and internet use (Table 3). Results revealed that younger individuals had significantly higher odds for internet use than older individuals (odds ratio [OR] 0.89, 95% CI 0.85-0.92; P<.001). In addition, gender was associated with internet use in older age: men were 2.8 times more likely to use internet than women. Furthermore, a higher educational level was significantly associated with internet use among older adults (OR 6.69, 95% CI 4.48-9.99; P<.001). Individuals who reported having a greater social network showed a higher chance for internet use than individuals with a smaller social network (OR 1.04, 95% CI 1.01-1.07; P=.01). Moreover, individuals with fewer depressive symptoms were 0.9 times more likely to use internet than those with more depressive symptoms. Participants with a better self-reported health-related quality of life had significantly higher odds for internet use than participants with lower self-reported health-related quality of life (OR 1.02, 95% CI 1.00-1.03; P=.006). Finally, the number of self-reported chronic diseases was associated with a higher likelihood of internet use among older adults: individuals who had a greater number of chronic diseases had 1.1 times the odds of reporting internet use than individuals with fewer chronic diseases, adjusting for all other sociodemographic and health-related covariates. No significant

associations between experience of loneliness, experience of social loss, marital status, and domicile and internet use were found. In total, this model explained between 29.3% (Cox and Snell) and 39.3% (Nagelkerke R^2) of the variance in internet use, and substantially improved model predictive power (from 56.8% to 73.8%).

Table 3. Results of binary logistic regression analysis for variables predicting internet use versus nonuse in older adults (N=937).

	Odds ratio (95% CI)	*P*-value	Wald
Age	0.888 (0.854-0.924)	<.001[h]	34.238
Gender			
Male	2.842 (2.017-4.002)	<.001[h]	35.711
Female	Reference		
Education[a]			
Low	Reference		
Middle	3.813 (2.582-5.630)	<.001[h]	45.295
High	6.691 (4.479-9.995)	<.001[h]	86.146
Marital status[b]			
Single	Reference		
Married, living in partnership	1.136 (0.614-2.100)	.685	0.165
Domicile			
Alone	Reference		
Living with others	1.354 (0.741-2.473)	.324	0.973
GDS[c]	0.888 (0.795-0.992)	.036[i]	4.407
LSNS[d]	1.037 (1.008-1.067)	.014[i]	6.101
Health-related quality of Life VAS EQ-5D-5L[e]	1.015 (1.004-1.025)	.006[j]	7.526
Chronic diseases[f]	1.122 (1.037-1.213)	.004[j]	8.301
UCLA Loneliness Scale[g]	0.905 (0.748-1.094)	.303	1.062
Experience of social loss			
Yes	0.855 (0.604-1.212)	.379	0.773
No	Reference		
Constant	407.376	.001	11.855
Nagelkerke R^2	0.393		
Log-likelihood	953.612		

[a]Educational classification according to the new CASMIN educational classification. Low: inadequately completed general education, general elementary education, basic vocational qualification or general elementary education, and vocational qualification; Middle: intermediate vocational qualification or intermediate general qualification and vocational qualification, intermediate general qualification, general maturity certificate, vocational maturity certificate/general maturity certificate, and vocational qualification; High: lower tertiary education-general diplomas/diplomas with vocational emphasis, higher tertiary education—lower level/higher level [41].

[b]Single: single, divorced, widowed; Married/living in partnership: married/with spouse, married/living apart).

[c]GDS: Geriatric Depression Scale; score ranges from 0 (no depression symptoms) to 15 (depression symptoms), cutoff: >4.

[d]LSNS: Lubben Social Network Scale; score ranges from 0 to 30, with higher values reflecting more social networks and more social support.

[e]VAS EQ-5D-5L: Visual Analogue Scale of the EuroQol 5-Dimensions 5-Level Questionnaire; score ranges from 0 (worst imaginable health state) to 100 (best imaginable health state).

[f]The sum score of chronic diseases ranges from 0 (no chronic conditions) to 22 (22 chronic conditions).

[g]UCLA: University of California, Los Angeles, Loneliness Scale; score ranges from 1 (hardly ever) to 3 (often).

[h]Significant on the level α=.001.

[i]Significant on the level α=.05.

[j]Significant on the level α=.01.

In order to test the association between health-related factors and the internet use for health purposes among internet users a binary logistic regression was conducted. Neither depressive symptoms (OR 1.10, 95% CI 0.962-1.260, P=.164) nor the number of chronic diseases (OR 1.03, 95% CI 0.932-1.132, P=.586) predicted internet search for health purposes.

Discussion

Principal Findings

The aim of this study was to investigate frequency of and factors associated with internet use among individuals aged 75 plus in Germany. Our results revealed that almost half of the 75- to 99-year old reported some internet use. This is comparable to the rates for adults aged 70 plus reported by a population survey on general internet use in Germany in 2016 [17]. Taking into account the trend of internet usage among this age group, there is evidence of a rapid increase of internet users in Germany [17]. Frees and Koch [17] showed that internet usage among this age group is on the rise (up to 65% in 2018), noting that older adults have the highest increase of daily internet use (from 25% in 2016 to about 40% in 2018). Our results confirm these findings—26.8% (261/999) of the 75-99-year olds used the internet daily. At the same time, we expanded existing data by examining factors related to internet use in older and oldest age. In sum, regression analyses showed that male gender, age, higher levels of education, a more extensive social network, higher level of health-related quality of life, fewer depressive symptoms, and a higher number of chronic diseases increased the likelihood of internet use in this age group.

Possible Explanations and Relation to Previous Studies

As for predisposing characteristics, the few studies that have focused on older adults (75+) [12,42-44] have shown that internet use in this age is associated with male gender, higher levels of education, being married, and reporting a better health status. In accordance therewith, our findings demonstrate that demographic and socioeconomic status variables are indeed significant predictors of internet use among this age group in Germany. Two of the strongest predictors are male gender and higher levels of education. van Deursen et al [45] showed in their (older adults) sample of internet users and nonusers in the Netherlands that women tend to not make use of the existing internet connection at home. The authors concluded that internet use among older adults seems to be a male-dominated activity. This may be due to the notion that ICT-related skills have been historically stereotypically perceived as more masculine (eg, males are good with technology). Among the older and oldest, such traditional stereotypes may play a more prominent role. However, as Baby Boomers get older, this may change [46], as younger generations of women are more familiar with using the internet and ICT (eg, smartphones).

As for the enabling factor social network/social support, previous studies that examined determinants of internet use among older adults have found that internet users also differ from nonusers with regard to their social embeddedness. Our findings are in accordance with previous research showing that internet users with a more extensive social network are more likely to use the internet [47]. However, due to the correlational nature of this study, no assertion can be made about causality. Therefore, the use of the internet might increase users' social network as measured in this study. Thus, there are other studies which highlight the potential of internet use to enhance or maintain social networks [2] and reduce feelings of loneliness [48] and social isolation [3]. Despite this, and in line with previous studies [12,45], our study found that support from partners, family members, or friends seems to effect the likelihood of going online.

As for need factors, existing studies show associations between internet use and physical/mental health and well-being [2,8,10,14] in this age group. Our findings also reflect this and show that the self-reported health-related quality of life increases the odds for internet use. In addition, a higher number of chronic medical conditions increased the likelihood of internet use in our multivariate analysis. A possible explanation might be the increased need for information about the disease and treatment options. Dumitru et al [19], for instance, noticed that using the internet for health-related purposes is very common (53.7%) and quickly increasing. Accordingly our findings show that more than half of the interviewees using the internet search on it for physical or psychological health information. Furthermore, findings from several studies show high rates of health-related internet use among those with multiple medical conditions or long-term illness [15,19]. However, this contrasts with our finding (ie, no link between internet use for health-related purposes and health-related factors such as the number of chronic medical conditions as well as depressive symptoms). Age seems to be negatively associated with health-related internet use [19,49,50]. This may be due to the fact that old and very old people with multiple medical conditions in Germany prefer other sources of information about health or illness, for example, direct face-to-face contacts with physicians and other health care professionals [19].

Common critical life events in older age are social loss experiences along with grief and bereavement, which are considered as risk factors for the development of mental health problems, especially depression and loneliness [51-53]. Whereas individuals with fewer depressive symptoms are more likely to use the internet in this study, no statistically significant effect of experiences of social loss or loneliness on an intensified use of internet among older adults was found. However, this does not necessarily mean older people would not benefit from ICT, for example, internet-based mental health interventions. Research has shown that eHealth (electronic health) interventions may significantly reduce mental health problems in older adults (eg, anxiety and depression [54-56]). Further research targeting this issue is currently being conducted [57].

Strengths and Limitations

The main strength of this study is the gathering of data from a population-based sample of the general German population aged 75 years and older that can be generalized to other populations in old age. Representativeness was ensured by using design and adjustment weighting methods of the data according to age, gender, and region [40] based on population statistics from German Federal Statistics [46]. However, it should be noted that representativeness may be limited due to the reduction

of the sample size for analysis and because only community-dwelling older adults with landline telephones, adequate hearing and speech comprehension, and without cognitive impairment were included in the sample. Apart from this, the study has further limitations. First, as the data from the telephone survey offer only a cross-sectional data set, no conclusion about long-term effects and causality can be drawn. Second, we aimed to incorporate many factors that may influence internet use, but, of course, it is likely that there were other unobserved variables at play, for instance, previous internet experiences or internet skills are associated with preferences in internet use in the older age group. Third, the survey is based on self-reported measures; therefore influences such as social desirability cannot be excluded. For more in-depth understanding of the use or nonuse of the internet and the underlying purposes qualitative research might be useful.

Conclusions

Today, public and private sectors commonly offer online services, which, in turn, have an influence on economic, cultural, and private life. ICT facilities include every day activities such as emailing, online banking, and information seeking, as well as increasingly eHealth services. For example, support via the internet, internet-based therapeutic interventions, and assistive technology are gaining in popularity. Hence, internet-based technologies have become a bigger part of the lives of older adults [58]. Therefore, there is a growing number of studies in this research field focusing specifically on older adults. The studies have identified factors associated with internet usage, as well as benefits and barriers. Moreover, eHealth interventions seem to be promising for promoting older people's well-being by fostering active aging or to helping the elderly stay independent as long as possible. Most of the current research has been mainly focused on younger people or has been limited to pilot studies [59]. Therefore, further research is needed, especially research identifying the types of older adults who would benefit the most from ICT use.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study sampling flowchart.
[PDF File (Adobe PDF File), 91 KB - jmir_v22i11e15543_app1.pdf]

References

1. Vrkljan B, Montpetit A, Naglie G, Rapoport M, Mazer B. Interventions that support major life transitions in older adulthood: a systematic review. Int Psychogeriatr 2019 Mar;31(3):393-415. [doi: 10.1017/S1041610218000972] [Medline: 29991365]
2. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
3. Chen YR, Schulz PJ. The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
4. Gatto SL, Tak SH. Computer, Internet, and E-mail Use Among Older Adults: Benefits and Barriers. Educational Gerontology 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
5. Näsi M, Räsänen P, Sarpila O. ICT activity in later life: Internet use and leisure activities amongst senior citizens in Finland. Eur J Ageing 2012 Jun;9(2):169-176 [FREE Full text] [doi: 10.1007/s10433-011-0210-8] [Medline: 28804417]
6. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de RSE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
7. Norris SL, High K, Gill TM, Hennessy S, Kutner JS, Reuben DB, et al. Health care for older Americans with multiple chronic conditions: a research agenda. J Am Geriatr Soc 2008 Jan;56(1):149-159. [doi: 10.1111/j.1532-5415.2007.01530.x] [Medline: 18047493]
8. Carpenter BD, Buday S. Computer use among older adults in a naturally occurring retirement community. Computers in Human Behavior 2007 Nov;23(6):3012-3024. [doi: 10.1016/j.chb.2006.08.015]
9. Edwards K, Duffy RM, Kelly BD. The Silver Surfer: Trends of Internet Usage in the Over 65 and the Potential Health Benefits. Ir Med J 2015 Jun;108(6):171-174. [Medline: 26182799]
10. Bujnowska-Fedak MM, Mastalerz-Migas A. Usage of medical internet and e-health services by the elderly. Adv Exp Med Biol 2015;834:75-80. [doi: 10.1007/5584_2014_74] [Medline: 25315621]

XSL•FO

RenderX

11.   White H, McConnell E, Clipp E, Bynum L, Teague C, Navas L, et al. Surfing the Net in Later Life: A Review of the Literature and Pilot Study of Computer Use and Quality of Life. J Appl Gerontol 2016 Jun 29;18(3):358-378. [doi: 10.1177/073346489901800306]

12.   Friemel TN. The digital divide has grown old: Determinants of a digital divide among seniors. 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]

13.   Hargittai E, Dobransky K. Old Dogs, New Clicks: Digital Inequality in Skills and Uses among Older Adults. CJC 2017 May 23;42(2):e1. [doi: 10.22230/cjc.2017v42n2a3176]

14.   Hunsaker A, Hargittai E. A review of Internet use among older adults. 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]

15.   Crabb RM, Rafie S, Weingardt KR. Health-related internet use in older primary care patients. Gerontology 2012;58(2):164-170. [doi: 10.1159/000329340] [Medline: 21734360]

16.   Tse MMY, Choi KCY, Leung RSW. E-health for older people: the use of technology in health promotion. Cyberpsychol Behav 2008 Aug;11(4):475-479. [doi: 10.1089/cpb.2007.0151] [Medline: 18721097]

17.   Frees B, Koch W. ARD/ZDF-Onlinestudie 2018: Zuwachs bei medialer internetnutzung und Kommunikation. Media Perspektiven. 2018. URL: https://www.ard-werbung.de/fileadmin/user_upload/media-perspektiven/pdf/2018/0918_Frees_Koch_2019-01-29.pdf [accessed 2020-07-13]

18.   Wirtschaftsrechnungen: Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien 2018. Statistisches Bundesamt. URL: https://tinyurl.com/y45a5u2g [accessed 2019-01-03]

19.   Dumitru RC, Bürkle T, Potapov S, Lausen B, Wiese B, Prokosch H. Use and perception of Internet for health related purposes in Germany: results of a national survey. Int J Public Health 2007 Oct;52(5):275-285. [doi: 10.1007/s00038-007-6067-0] [Medline: 30311153]

20.   Hüfken V, Deutschmann M, Baehring T, Scherbaum W. Verbreitung und Sozialprofil der gesundheitsthemenbezogenen Internetnutzung: Ergebnisse einer bundesweiten Telefonumfrage. Soz -Prventivmed 2004 Dec;49(6):381-390. [doi: 10.1007/s00038-004-3149-0]

21.   Anderson R. A behavioral model of families' use of health services. In: Research Series No 25. Chicago: IL: Center for Health Administration Studies, University of Chicago; 1968:xi-111.

22.   Stein J, Liegert P, Dorow M, König H, Riedel-Heller SG. Unmet health care needs in old age and their association with depression – results of a population-representative survey. Journal of Affective Disorders 2019 Feb;245:998-1006. [doi: 10.1016/j.jad.2018.11.096]

23.   Bickel H, Ambros C, Schulze Beerhorst B. Retrospektive Demenzeinschätzung durch Informantenbefragungen. Ergebnisse zur Validität und zur Inter-Rater-Reliabilität. Zeitschrift für Gerontopsychologie und -psychiatrie 1993;6:175-182.

24.   Brooke P, Bullock R. Validation of a 6 item cognitive impairment test with a view to primary care usage. Int J Geriatr Psychiatry 1999 Nov;14(11):936-940. [Medline: 10556864]

25.   Gauggel S, Birkner B. Validität und Reliabilität einer deutschen Version der Geriatrischen Depressionsskala (GDS). Zeitschrift für Klinische Psychologie und Psychotherapie 1999 Jan;28(1):18-27. [doi: 10.1026//0084-5345.28.1.18]

26.   Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriat Psychiatry 1999 Oct;14(10):858-865. [doi: 10.1002/(sici)1099-1166(199910)14:10<858::aid-gps35>3.0.co;2-8]

27.   Allen J, Annells M. A literature review of the application of the Geriatric Depression Scale, Depression Anxiety Stress Scales and Post-traumatic Stress Disorder Checklist to community nursing cohorts. J Clin Nurs 2009 May;18(7):949-959. [doi: 10.1111/j.1365-2702.2008.02731.x] [Medline: 19284431]

28.   Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]

29.   Holland R, Smith RD, Harvey I, Swift L, Lenaghan E. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. Health Econ 2004 Aug;13(8):793-805. [doi: 10.1002/hec.858] [Medline: 15322991]

30.   Hinz A, Kohlmann T, Stöbel-Richter Y, Zenger M, Brähler E. The quality of life questionnaire EQ-5D-5L: psychometric properties and normative values for the general German population. Qual Life Res 2014 Mar;23(2):443-447. [doi: 10.1007/s11136-013-0498-2] [Medline: 23921597]

31.   Hunger M, Sabariego C, Stollenwerk B, Cieza A, Leidl R. Validity, reliability and responsiveness of the EQ-5D in German stroke patients undergoing rehabilitation. Qual Life Res 2012 Sep;21(7):1205-1216. [doi: 10.1007/s11136-011-0024-3] [Medline: 21971874]

32.   Stark RG, Reitmeir P, Leidl R, König H. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. Inflamm Bowel Dis 2010 Jan;16(1):42-51. [doi: 10.1002/ibd.20989] [Medline: 19475674]

33.   Sonntag M, Konnopka A, Leichsenring F, Salzer S, Beutel ME, Herpertz S, et al. Reliability, validity and responsiveness of the EQ-5D in assessing and valuing health status in patients with social phobia. Health Qual Life Outcomes 2013 Dec 23;11:215 [FREE Full text] [doi: 10.1186/1477-7525-11-215] [Medline: 24365384]

XSL•FO

RenderX

34. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]

35. Holmes TH, Rahe RH. The social readjustment rating scale. Journal of Psychosomatic Research 1967 Aug;11(2):213-218. [doi: 10.1016/0022-3999(67)90010-4]

36. Rahe RH. Epidemiological studies of life change and illness. Int J Psychiatry Med 1975;6(1-2):133-146. [doi: 10.2190/JGRJ-KUMG-GKKA-HBGE] [Medline: 773851]

37. Brown GW, Harris T. Social Origins of Depression: A Study of Psychiatric Disorder in Women. London: Tavistock; 1978.

38. Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance of an Abbreviated Version of the Lubben Social Network Scale Among Three European Community-Dwelling Older Adult Populations. The Gerontologist 2006 Aug 01;46(4):503-513. [doi: 10.1093/geront/46.4.503]

39. Zensus 2011. Statistisches Bundesamt. 2011. URL: https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6 [accessed 2019-01-03]

40. Gabler S, Kolb JP, Sand M, Zins S. Weighting. GESIS Survey Guidelines. 2016. URL: https://www.gesis.org/fileadmin/upload/SDMwiki/Gabler_et_al_Weighting.pdf [accessed 2019-01-03]

41. Brauns H, Steinmann S. Educational reform in France, West-Germany and the United Kingdom: updating the CASMIN educational classification. ZUMA Nachrichten 1999:7-44 [FREE Full text]

42. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]

43. Peacock SE, Künemund H. Senior citizens and Internet technology: Reasons and correlates of access versus non-access in a European comparative perspective. Eur J Ageing 2007 Dec;4(4):191-200 [FREE Full text] [doi: 10.1007/s10433-007-0067-z] [Medline: 28794788]

44. Selwyn N. The information aged: A qualitative study of older adults' use of information and communications technology. Journal of Aging Studies 2004 Nov;18(4):369-384. [doi: 10.1016/j.jaging.2004.06.008]

45. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. European Journal of Communication 2015 Apr;30(2):171-187. [doi: 10.1177/0267323115578059]

46. Older people in Germany and the EU 2016. Wiesbaden, Germany: Federal Statistical Office of Germany URL: https://www.destatis.de/EN/Publications/Specialized/Population/BrochureOlderPeopleEU0010021169004.pdf?__blob=publicationFile [accessed 2019-01-03]

47. Nahm ES, Resnick B, Mills ME. A Model of Computer-Mediated Social Support Among Older Adults. AMIA Annu Symp Proc 2003:948 [FREE Full text] [Medline: 14728453]

48. Pettigrew S, Roberts M. Addressing loneliness in later life. Aging Ment Health 2008 May;12(3):302-309. [doi: 10.1080/13607860802121084] [Medline: 18728942]

49. Hong YA, Cho J. Has the Digital Health Divide Widened? Trends of Health-Related Internet Use Among Older Adults From 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]

50. Levine DM, Lipsitz SR, Linder JA. Trends in Seniors' Use of Digital Health Technology in the United States, 2011-2014. JAMA 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]

51. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]

52. Förster F, Stein J, Löbner M, Pabst A, Angermeyer MC, König H, et al. Loss experiences in old age and their impact on the social network and depression- results of the Leipzig Longitudinal Study of the Aged (LEILA 75+). J Affect Disord 2018 Dec 01;241:94-102. [doi: 10.1016/j.jad.2018.07.070] [Medline: 30107351]

53. Sikorski C, Luppa M, Heser K, Ernst A, Lange C, Werle J, AgeCoDe Study Group. The role of spousal loss in the development of depressive symptoms in the elderly - implications for diagnostic systems. J Affect Disord 2014 Jun;161:97-103. [doi: 10.1016/j.jad.2014.02.033] [Medline: 24751315]

54. Jones SL, Hadjistavropoulos HD, Soucy JN. A randomized controlled trial of guided internet-delivered cognitive behaviour therapy for older adults with generalized anxiety. J Anxiety Disord 2016 Jan;37:1-9. [doi: 10.1016/j.janxdis.2015.10.006] [Medline: 26561733]

55. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]

56. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided. self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]

57. Brodbeck J, Berger T, Znoj HJ. An internet-based self-help intervention for older adults after marital bereavement, separation or divorce: study protocol for a randomized controlled trial. Trials 2017 Dec 13;18(1):21 [FREE Full text] [doi: 10.1186/s13063-016-1759-5] [Medline: 28086836]

XSL•FO

RenderX

58.   Wagner N, Hassanein K, Head M. Computer use by older adults: A multi-disciplinary review. Computers in Human Behavior 2010 Sep;26(5):870-882. [doi: 10.1016/j.chb.2010.03.029]

59.   Preschl B, Forstmeier S, Maercker A, Wagner B. E-health interventions for depression, anxiety disorders, dementia, and other disorders in old age: A review. Journal of Cyber Therapy and Rehabilitation 2011;4(4):371-385. [doi: 10.5167/uzh-67320]

Abbreviations

AMD: Association for German Market and Social Research Agency
CASMIN: CASMIN Educational Classification in International Comparative Research
DFG: German Research Foundation
GDS: Geriatric Depression Scale
ICT: Information and Communication Technology
LLL: Leipziger Lebensereignis-Liste
LSNS: Lubben Social Network Scale
SPSS: Statistical Package for the Social Sciences
UCLA: University of California, Los Angeles, Loneliness Scale

Original Paper

Exploring Novel Funding Strategies for Innovative Medical Research: The HORAO Crowdfunding Campaign

Philippe Schucht[1*], MD; Diana M Roccaro-Waldmeyer[2,3*], PhD; Michael Murek[1], MD; Irena Zubak[1], MD; Johannes Goldberg[1], MD; Stephanie Falk[1], MA; Fried-Michael Dahlweid[2,4], MD, PhD; Andreas Raabe[1], MD

[1]Department of Neurosurgery, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[2]Insel Data Science Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[3]Directorate of Technology and Innovation, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[4]DXC Technology, Tysons, VA, United States

[*]these authors contributed equally

Corresponding Author:
Philippe Schucht, MD
Department of Neurosurgery
Inselspital, Bern University Hospital
University of Bern
Freiburgstrasse
Bern, 3010
Switzerland
Phone: 41 31 664 02 21
Email: philippe.schucht@insel.ch

Abstract

Background: The rise of the internet and social media has boosted online crowdfunding as a novel strategy to raise funds for kick-starting projects, but it is rarely used in science.

Objective: We report on an online crowdfunding campaign launched in the context of the neuroscience project HORAO. The aim of HORAO was to develop a noninvasive real-time method to visualize neuronal fiber tracts during brain surgery in order to better delineate tumors and to identify crucial cerebral landmarks. The revenue from the crowdfunding campaign was to be used to sponsor a crowdsourcing campaign for the HORAO project.

Methods: We ran a 7-week reward-based crowdfunding campaign on a national crowdfunding platform, offering optional material and experiential rewards in return for a contribution toward raising our target of Swiss francs (CHF) 50,000 in financial support (roughly equivalent to US $50,000 at the time of the campaign). We used various owned media (websites and social media), as well as earned media (press releases and news articles) to raise awareness about our project.

Results: The production of an explanatory video took 60 hours, and 31 posts were published on social media (Facebook, Instagram, and Twitter). The campaign raised a total of CHF 69,109. Approximately half of all donations came from donors who forwent a reward (CHF 28,786, 48.74%); the other half came from donors who chose experiential and material rewards in similar proportions (CHF 14,958, 25.33% and CHF 15,315.69, 25.93%, respectively). Of those with an identifiable relationship to the crowdfunding team, patients and their relatives contributed the largest sum (CHF 17,820, 30.17%), followed by friends and family (CHF 9288, 15.73%) and work colleagues (CHF 6028, 10.21%), while 43.89% of funds came from donors who were either anonymous or had an unknown relationship to the crowdfunding team. Patients and their relatives made the largest donations, with a median value of CHF 200 (IQR 90).

Conclusions: Crowdfunding proved to be a successful strategy to fund a neuroscience project and to raise awareness of a specific clinical problem. Focusing on potential donors with a personal interest in the issue, such as patients and their relatives in our project, is likely to increase funding success. Compared with traditional grant applications, new skills are needed to explain medical challenges to the crowd through video messages and social media.

KEYWORDS

Introduction

In 1897, the French scientist Gaston Contremoulins was receiving insufficient support for his work from governmental agencies and therefore turned to the crowd via the popular French newspaper "*Le temps*" to ask for financial help. In what was essentially the first crowdfunded neuroscience project, he and his team raised enough capital to perform the first stereotactic surgery in humans [1]. More than a century later, in response to the financial crisis of 2007 and 2008, and thanks to the rise of the internet and e-commerce, crowdfunding has emerged as an alternative method to raise funds and has been gaining popularity ever since [2,3]. Especially for high-risk projects, which are rarely supported by traditional funding agencies [4,5], crowdfunding represents a potentially more promising alternative. Crowdfunding, a term first coined in 2007 [6], can be defined as an innovative method of fundraising for a project or business, which typically involves a large number of rather small contributions from individuals following an open call through the internet [7,8]. In the health care domain, the competition for medical research funding from government sources, such as the National Institutes of Health, is continuously increasing [4,5,9], making crowdfunding an attractive alternative for clinical and scientific research funding. Unlike traditional funding mechanisms, it does not involve a rigorous peer-review process, but instead offers a way to reach more diverse audiences and to raise public awareness about scientific problems [5,10,11].

From an economic perspective, the following four different business models of crowdfunding can be distinguished, depending on what donors receive in return for their financial contributions: donation-based crowdfunding (no return on investment expected), reward-based crowdfunding (optional nonmonetary returns), lending-based crowdfunding (return of the funds, possibly with interest), and equity-based crowdfunding (future profit of the venture is shared) [7-9,12]. Crowdfunding campaigns aiming to foster medical research are often reward-based [9]. Although reward-based crowdfunding can involve prepurchasing a newly developed product, it is understood here as donating in return for nonfinancial potentially experiential rewards of little economic, but rather symbolic, nontradable value [8,9].

Since the model requires a two-sided market, almost all instances of internet-based crowdfunding fundamentally depend on crowdfunding platforms that act as intermediaries linking fundraisers to potential funders [7,13]. These platforms are essential to create a trustworthy environment in which donors feel secure enough to exchange money for rewards [14]. They also provide a space to describe the project, facilitate communication with potential donors, and help to run the campaign while ensuring standardized processes [7].

During brain tumor surgery, distinguishing tumor tissue from the surrounding healthy tissue remains a challenge. An ideal technology would noninvasively and reproducibly visualize tumor tissue, without time loss in real-time and without harming the surrounding brain. Although technologies, such as 5-aminolevulinic acid fluorescence, are useful to visualize a subset of tumors during surgery [15], as yet it has not been possible to develop a technology applicable to all intrinsic tumors of the brain. We therefore decided to shift the focus from direct identification of tumor tissue to visualization of the microstructure of the brain. Fiber tracts are a hallmark of the white matter of the brain and cannot be seen in tumor tissue. A technology that identifies fiber tracts during surgery would thus allow the surgeon to differentiate between tumor tissue and healthy brain. In order to catalyze interdisciplinary research in this field and to foster the development of an optical instrument that interfaces with current state-of-the-art microscopes to improve the visualization of brain tumor boundaries, we initiated a global, crowdsourced, scientific competition. The aim of the crowdfunding campaign described here was to raise the prize money as an incentive for the above-mentioned crowdsourcing competition.

To our knowledge, this is the first report on an internet-based crowdfunding campaign for clinical neuroscience to be published in a medical journal. Here, we describe our experience with crowdfunding to finance a neuroscience project and analyze its strengths and the challenges.

Methods

Crowdfunding Campaign

The campaign was launched on the Science Booster Channel of wemakeit, a leading Swiss crowdfunding platform, on August 3, 2017 [16] (Figure 1). The funding target was set at 50,000 Swiss francs (CHF; roughly equivalent to US $50,000 at the time of the campaign), and the duration of the campaign was 47 days. Depending on how much they donated, donors could choose from 12 different rewards or decide to forgo the reward. Material rewards included project t-shirts, thank-you cards and plasters designed explicitly for this campaign, a bestseller novel signed by the team of physicians, and a unique work of art created exclusively for the purpose of this campaign that was donated by the artist. Experiences on offer included a visit to an artist's studio, an invitation to the scientific conference to be held at the end of the crowdsourcing competition, a guided tour of the neurosurgical department, neurosurgical training in the skills lab, a one-day visit to the neurosurgical department with the opportunity to shadow physicians, and a dinner with the team of physicians. Donors were classified according to their relationship to the crowdfunding team as (1) friends and family members, (2) work colleagues, (3) patients (and their relatives), (4) unknown, and (5) anonymous.

Figure 1. Screenshot taken from the online donation platform at the time of completion of the HORAO campaign.

Media Activity

Owned (websites and social media) and earned media (press releases and news articles) were used to raise public awareness about the campaign. A website was created specifically for the purpose of this crowdfunding and the subsequent crowdsourcing campaign [17]. A 3-minute explanatory video was produced and published on wemakeit, as well as on the YouTube channel of our hospital group (Multimedia Appendix 1) [18]. The video featured three brain tumor patients and had an overall production time of 60 hours. Involvement of social media required daily attention. A total of 31 posts and tweets were published on Facebook, Instagram, and Twitter during the course of the campaign. Two press releases on August 03 and September 19, 2017, resulted in the publication of 14 media reports (print, online, and radio) at the beginning (eight reports on August 03 and six reports on August 04), and nine media reports (print and online) at the end of the campaign. During the campaign,

a report was printed in the largest Swiss newspaper *20 Minuten* (on August 08, 2017), supplemented by a sponsored short article that continued to be featured in the electronic version of the newspaper throughout the week of August 18 to August 24 (a pro-bono donation made by the news portal). A contribution to the newsletter of the European Association of Neurosurgical Societies was also submitted. Finally, regular updates were published on the websites of our hospital group, our university hospital, our neurosurgical department, and our cancer center.

Statistical Analysis

IBM SPSS Statistics Version 25 (IBM Corp) was used for all statistical analyses and for the creation of the graphs. A Kruskal-Wallis H test was performed to compare total donations per day across the seven campaign weeks, as well as to compare the size of individual donations between donor groups with different relationships to the crowdfunding team. Bonferroni correction was used to adjust for multiple post-hoc comparisons.

A Mann-Whitney *U* test was performed to investigate the effect of publicity in the newspaper *20 Minuten* (August 08 to August 17 vs August 19 to August 25). A Pearson chi-squared test was performed to assess the correlation between reward type and type of relationship, while Goodman and Kruskal tau served to determine the extent of the bidirectional interdependence. Statistical significance was set at $P<.05$.

Results

On campaign day 30 (August 31, 2017), the funding goal of CHF 50,000 was reached (Figure 2). The crowdfunding campaign ended after the predefined duration of 47 days. A total of CHF 69,109 (contributed by 235 donors) was raised, exceeding the target of CHF 50,000 by 38.22%. Our expenditures were as follows: CHF 6911 for the crowdfunding platform (10.00% of the total revenue), CHF 4839 for the production of the video, CHF 1736 for the purchase of the material rewards, and CHF 280 for web services, resulting in a total revenue of CHF 55,443. Of this total, CHF 50,000 was used as prize money for the crowdsourcing campaign, and the rest (CHF 5,443) was used for the organization of the final crowdsourcing conference, during which the finalists described their solution to the public and competed for the prize.

Figure 2. Cumulative sum of daily donations across the 7 weeks of the HORAO campaign. The dotted horizontal line indicates the funding goal of Swiss francs (CHF) 50,000 was reached on day 30, triggering the contribution of an additional CHF 10,000 by Gebert Rüf Stiftung (asterisk). The dashed vertical lines delimit the 7-day period during which an article on the project was featured in the electronic version of the newspaper *20 Minuten* (days 16-22 of the campaign).
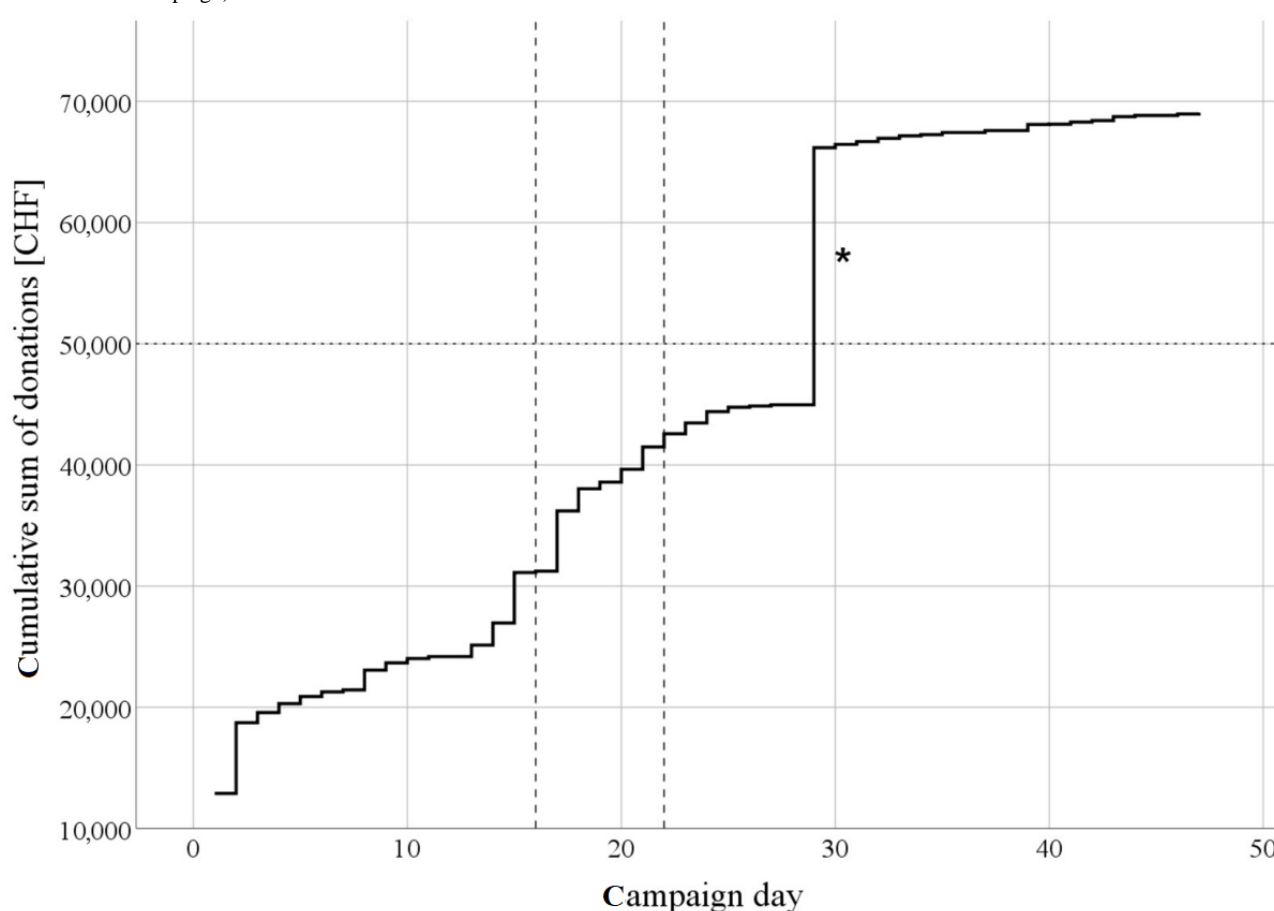
Our crowdfunding campaign received CHF 10,000 in additional support from the Science Booster Channel of wemakeit thanks to a generous grant from Gebert Rüf Stiftung (Projekt-Nr. GRS-040/15) upon reaching the predetermined target. This additional prize money has been excluded from the analysis of the funding campaign (ie, 100% always refers to CHF 59,109).

Donations per day tended to be higher in the first four than in the last two weeks of the campaign (Kruskal-Wallis H test: $P=.03$). However, after adjusting for multiple comparisons, none of the differences remained statistically significant. The first week was the most successful (during which 35.84% [CHF 21,185] of the total sum of CHF 59,109 was raised), followed by the fifth week (CHF 12,171, 20.59%) (Table 1). Twenty-two percent of the total sum was raised on day 2 alone, followed by 19% on day 30 and 10% on day 3. On five of the 48 days, no donations were made, and four out of these five days were in the final three weeks. During the 7 days when our project was featured in the print version (August 19) and in the electronic version of the newspaper *20 Minuten* (August 19 to August 25), the sum of total donations was approximately 50% higher (CHF 4173) than during the preceding 7 days (August 11 to August 17; CHF 12,229 vs CHF 8056). Additionally, the number of individual donations per day was higher in the second week (median 11, IQR 8 vs median 3, IQR 4; $P=.01$), although the value of total donations per day did not significantly differ between the two weeks (median CHF 1090, IQR 4426 vs median CHF 600, IQR 1660; $P=.17$). In addition, the contributions from

both anonymous donors and those whose relationship to the crowdfunding team was unknown peaked during these 7 days (August 19 to August 25; 55% of total donations). These donors had accounted for only 28% of donations in the 7 days before the publicity in *20 Minuten*. The online version of the article generated a total of 25,911 views (869–5595 views per day). There was a strong correlation between the number of views per day and the total number of donations made on that day (r=0.94, P<.001) (Figure 3). On average, 368 clicks generated one extra donation (average donation: CHF 61). The news portal offered this online exposure free of charge as a donation, and its commercial value was estimated at approximately CHF 20,000.

Table 1. Descriptive statistics of donations per day over the 7 weeks.

Week	Median per day (CHF[a])	Mean per day (CHF)	Maximum per day (CHF)	Sum per week (CHF) (N=59,109), value (%)
1 (August 03 to August 09)	661	2648	12,598	21,185 (35.84)
2 (August 10 to August 16)	570	610	1631	4272 (7.23)
3 (August 17 to August 23)	560	1432	4970	10,025 (16.96)
4 (August 24 to August 30)	937	1318	4097	9228 (15.61)
5 (August 31 to September 06)	230	1739	11,211	12,171 (20.59)
6 (September 07 to September 13)	125	183	500	1284 (2.17)
7 (September 14 to September 18)	100	129	320	644 (1.09)

[a]CHF: Swiss francs.

Figure 3. Correlation between the number of clicks per day and the total number of donations made on that day, for the 7 days during which HORAO was featured in the electronic version of the newspaper *20 Minuten* as well as the 7 days before (r=0.94, P<.001). Points represent individual days. On average, 368 clicks triggered one extra donation (average value of donation: Swiss francs [CHF] 61.40), corresponding to a mean donation of CHF 0.17 per click.
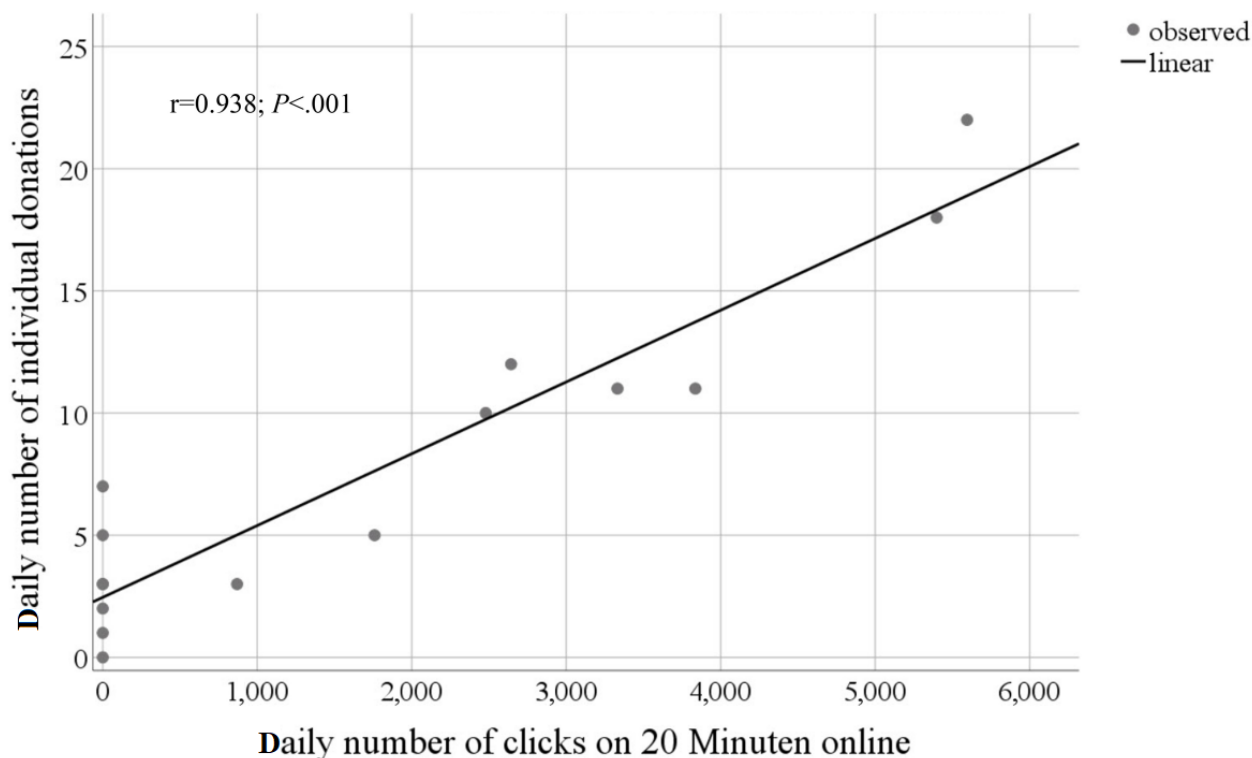
Of the groups with a known relationship to the crowdfunding team, the one comprising patients (and their relatives) made the largest contribution to the funds (30.17%), followed by the groups of friends and family members, anonymous donors, and

work colleagues (Table 2). Donors whose relationship to the crowdfunding team was unknown contributed the largest proportion of all funds raised (32.04%).

The median donation value per donor was significantly higher for the group of patients (and their relatives) (CHF 200, IQR 90) than for anonymous donors (CHF 50, IQR 80; P=.001) and for donors whose relationship to the crowdfunding team was unknown (CHF 87, IQR 140; P<.001).

Similar percentages of the funding came from donors who selected experiences (25.33%) and those who selected material rewards (25.93%), but most funding (48.74%) came from donors who forwent a reward (Table 2). In terms of the 12 individual rewards offered, most funding came from donors who selected the neurosurgical training in the skills lab (total CHF 7215), followed by those who selected the t-shirt (total CHF 4640), the work of art (CHF 4000), and the invitation to the scientific conference to be held at the end of the crowdsourcing competition (total CHF 3750). As far as it is possible to tell, donations came from 11 countries spread across three continents (Europe, North America, and Asia) (Table 2). The vast majority of funds came from donors based in Switzerland (82.33%), while the geographic location of 29.1% of donors (68/234) contributing 12.33% of the total donation sum was unknown.

Table 2. Characteristics of donations.

Characteristic	Donation (CHF[a])	Value, %
Relationship		
Patients and their relatives	17,820.00	30.17%
Unknown	18,922.69	32.04%
Friends	9288.00	15.73%
Colleagues	6028.00	10.21%
Anonymous	7001.00	11.85%
Total	59,059.69	100.00%
Type of reward		
Material	15,315.69	25.93%
Experiences	14,958.00	25.33%
Gift declined	28,786.00	48.74%
Total	59059.69	100.00%
Country of origin		
Switzerland	48,624.00	82.33%
Great Britain	1047.00	1.77%
Singapore	876.00	1.48%
Germany	272.69	0.46%
United States	259.00	0.44%
Greece	250.00	0.42%
Luxembourg	200.00	0.34%
Lebanon	160.00	0.27%
Austria	50.00	0.08%
Poland	20.00	0.03%
Liechtenstein	20.00	0.03%
Unknown	7281.00	12.33%
Total	59,059.69	100.00%

[a]CHF: Swiss francs.

Discussion

Principal Results

Can neuroscientists use internet-based crowdfunding to kick-start new projects? The HORAO (from ancient Greek "to see with one`s mind") campaign was able to reach its funding goal of CHF 50,000 within a month. Ultimately, we received 138% of the target (CHF 69,109) from 235 donors. This success is somewhat contradictory to the experiences reported in the scientific literature suggesting that crowdfunding can hardly yield sums larger than US $10,000 [4,10,11] and that similar to traditional funding agencies, donors are hesitant about

supporting innovative projects [19]. What are the specific aspects of our campaign that could have contributed to its success?

Contributors

The principle of crowdfunding consists of a large number of individuals contributing generally small amounts each. In our campaign, the group of donors whose relationship to the crowdfunding team was unknown contributed more funding than any of the other groups. However, the efforts made to reach out to the crowd were substantial. For instance, the financial cost of the sponsored 1-week online newspaper campaign would likely have exceeded the additional donations it brought in, if we had to pay for it. Hence, while we believe that reaching out to the crowd through news media and online portals will help to attain the funding goals, it is unlikely to be profitable if market prices have to be paid.

Patients and their relatives contributed the largest amount of funds. Given that patients and their relatives might one day themselves benefit from the success of our campaign, this finding comes as no surprise. Nevertheless, our results contradict a conclusion from a crowdfunding study on more general product marketing that compared the effectiveness of different types of appeal in raising money for product development [20]. This study found that product benefits to donors do not seem to be the most effective strategy. The incentive to donate money seems to obey different laws when it comes to life-threatening diseases such as cancer. Hence, we believe that identifying potential donors with a specific personal interest in the field of research is a key step toward ensuring the success of crowdfunding campaigns for medical research. Contacting specific interest groups, such as patient organizations with an interest in brain tumor treatment, might have attracted further donations. However, in our campaign, we refrained from contacting patients and their relatives directly and from encouraging them to donate in any way to avoid raising ethical concerns. Patients might otherwise get the false impression that failure to donate could entail negative consequences for their treatment.

Most of the funds raised during the first 2 weeks came from our existing network, consisting of not only friends, family, and work colleagues, but also patients and their relatives. However, unknown donors also started to contribute from the very first week. Our findings thus agree with the general observation that the initial donations in crowdfunding campaigns typically stem from the project creators' own social networks [21,22]. However, our findings differ somewhat from those of an eHealth study on congenital heart disease, where unknown donors only started to contribute after observing some initial funding success [21]. We believe that the continuous preliminary campaign with attention-grabbing videos on social media in the weeks preceding the start of the crowdfunding campaign was responsible for the early buy-in by unknown donors. Hence, while it takes more effort to convince unknown donors to contribute, preliminary activities on social media might prime these potential donors to recognize the quality of the campaign, prompting them to donate earlier. The fact that at least 82% of our total funding came from donors located in Switzerland

reconfirms the "home bias," which is characteristic of crowdfunding transactions, that is, the crowd tends to support funders located in proximity to themselves [13]. Overall, our results confirm the importance of a large social network, which in this case included not only family and friends, but also patients, for success in obtaining funding.

Timeline of Contributions and the Influence of the Media

Donations did not follow a linear pattern throughout the campaign. Instead, we observed a pronounced peak at the beginning, which is best explained by the actions of donors from our own network who had been primed for the campaign and donated once it started. In addition, we observed an increase in donation activity once the funding goal came within reach, presumably because this led hesitant donors to feel confident about the success of the campaign. The timeline of our donations replicates the U-shaped pattern according to which donors support campaigns preferably at their start and when funding approaches the target sum [23]. Hence, the dry period after the initial peak in funding requires additional effort. Our strategy of bridging the dry period with a second media campaign was successful to a certain extent.

It is difficult to assess the effect of the 14 media reports published on August 03 and August 04, 2017, since these coincided with the launch of the HORAO campaign. The article that appeared in both print and online versions during week three of the campaign resulted in a peak in contributions from both unknown and anonymous donors lasting several days, and thus, undoubtedly revived contributions from donors outside our own social network. In the literature, social media activity is reported to correlate with fundraising success, but the strength of this effect seems to depend on the type of campaign (creative, social, or entrepreneurial) [24]. To summarize, preparation of material to maintain a media presence is time-consuming and requires extra effort (eg, to create explanatory videos), but can help to bridge the slack period.

Since more than half of our funding goal (CHF 26,000) was contributed by the five largest donations and these were made by individuals from five different groups (all but the "unknown"), it is not possible to conclude that any one group would have been more critical to the success of our campaign than another. Most likely, our results demonstrate that the personal experience of affected patients and their relatives, in combination with their personal relationship to a physician launching a crowdfunding campaign, can be a powerful incentive to donate.

A study investigating the strategies and communication tools of reward-based crowdfunding campaigns identified the following three different paths to funding success, dependent on the type of project: "communicator," "networker," and "self-runner" strategies [12]. Self-runner projects catch the attention of the crowd, including the media, all by themselves, so rewards are hardly needed to attract backers. By contrast, communicator projects are rather weak by themselves, and their success depends on concerted efforts in online marketing and public relations as well as well-chosen rewards. Networker projects rely more on the personal network of the funders than

communicator projects do, and rewards are essential to attract the personal network, while the information will then automatically flow to the general public, making web presence superfluous. Although a sufficient number of donors is a prerequisite for funding success independent of the path, social media activity is critical only in two of the three paths (communicator and networker). In these cases, the attention the crowd can pay to the projects is not sufficient at the outset, so regularly updated information is needed to maintain their interest. Moreover, rewards and web presence are thought to substitute for one another. It is conceivable that for patients and their relatives, our campaign was a "self-runner" project, whereas networking and regular communication were critical in attracting donors who did not have a personal link to the problem at stake.

Selecting the Right Rewards

It is difficult to assess whether we selected the right rewards. None of the rewards clearly stood out in terms of funding success, and both experiential and material rewards contributed similarly to overall funding success. When choosing the number of rewards, we followed the recommendation in the literature, which had shown an inverted U-shaped relationship between the number of rewards offered in a crowdfunding campaign and the number of donors attracted (a moderate number of rewards is preferable to a lengthy list) [25]. Every third donor (35%) chose to forgo the reward, and half of our total funding was contributed by donors who did not claim any rewards. Our results seem to emphasize the importance of donors who do not require any incentives at all, presumably because our project represented an issue that is close to their heart.

Crowdfunding Versus Traditional Funding

We assumed that because our project is largely application-oriented, it is not the type of project favored by traditional funding instruments. Moreover, as a strategically innovative and thus high-risk project, the chances of obtaining traditional funding are low [4]. For these reasons, we decided to address the crowd with our request. The ever-increasing competition for traditional (typically governmental) sources of project funding is the primary reason why project creators turn toward crowdfunding as an alternative way to obtain financing [4,11]. Crowdfunding offers the potential to reach a much broader and more diverse audience than traditional funding applications [10]. Besides the obvious benefit of obtaining funds, involving the crowd has the potential to bring more transparency into the mechanisms of science funding, as well as ideally building a lasting community and raising awareness [10].

New Skills Needed

One of the main challenges of this project was to explain a specific scientific problem to the crowd, which requires a different communication strategy and different presentation skills from those needed to write grant applications for peer-reviewed funding [4]. Potential donors are likely to spend less time studying a project description and to have less pre-existing knowledge about this research area than peer reviewers, resulting in knowledge asymmetry [26]. Hence, the project description must be easily comprehensible and presented concisely in layman's terms. The ability of scientists to engage with the crowd and build an audience around their project is one of the most quoted factors in successful scientific crowdfunding [4,26]. In addition to more factual written descriptions, short videos provide the narrative that captures the attention of a potential donor and helps to evoke emotions that trigger donations [10]. Coaching by crowdfunding professionals from the commercial crowdfunding platform (wemakeit) and by other scientists who had run crowdfunding campaigns was crucial in enabling us to identify and acquire the necessary skill set.

Outlook

Research studies have attempted to establish crowdfunding as one form of entrepreneurial, technology, or science-related financing; however, little is known about the sustainability of crowdfunding projects. Crowdfunding can be split into the following two stages: funding and implementation [27]. A campaign's success in reaching its funding goal does not necessarily guarantee success in project implementation. While in traditional funding processes, peer review is supposed to channel funding to the most promising projects, concerns have been raised about the scientific quality and sustainability of successfully crowdfunded science projects [5]. On the other hand, projects supported by the crowd might contribute to scientific outcomes more tailored to the needs of the general public than projects selected by the inherently biased opinion of a small number of peer reviewers. As opposed to the extensively studied funding stage, very little research has been published on the implementation of crowdfunding projects [27]. Further studies are needed to investigate the performance of scientific and technological ventures after successful crowdfunding, as well as the impact of crowdfunding on science in general.

Limitations

The external validity of our results may be limited. Campaign success depends on the size of the team's social network and on the specific goal. Cancer-related issues may trigger stronger emotions than other neuroscientific topics, which might have contributed to the success of the present campaign. Some benefits, such as promotion on the news portal and support from the foundation, are because of the project's novelty. In addition, the success of our campaign should not be seen as proof of the scientific value of our endeavor.

XSL•FO
RenderX

video teasers largely pro-bono, to Loretta Shebaya and her team at Paper View Design for creative support, and to the artist Martina von Meyenburg for donating one of her works. We thank Stefan Halter and 20 Minuten for promotion of our project.

Multimedia Appendix 1

Promotional video from the crowdfunding website.

[MP4 File (MP4 Video), 55878 KB - jmir_v22i11e19715_app1.mp4]

References

1. Bourdillon P, Apra C, Lévêque M. First clinical use of stereotaxy in humans: the key role of x-ray localization discovered by Gaston Contremoulins. J Neurosurg 2018 Mar;128(3):932-937. [doi: 10.3171/2016.11.JNS161417] [Medline: 28498054]

2. Dix A, Luzar C. Current Market Dynamics. In: Dresner S, editor. Crowdfunding: A Guide to Raising Capital on the Internet. New Jersey, United States: John Wiley & Sons; Mar 2014:47-80.

3. Janku J, Kucerová Z. Successful crowdfunding campaigns: The role of project specifics, competition and founders' experience. Czech Journal of Economics and Finance 2018;68(4):351-373.

4. Byrnes JE, Ranganathan J, Walker BL, Faulkes Z. To Crowdfund Research, Scientists Must Build an Audience for Their Work. PLoS One 2014;9(12):e110329 [FREE Full text] [doi: 10.1371/journal.pone.0110329] [Medline: 25494306]

5. Siva N. Crowdfunding for medical research picks up pace. The Lancet 2014 Sep;384(9948):1085-1086. [doi: 10.1016/s0140-6736(14)61661-5]

6. Herther N. Crowdfunding: Research and Academic Fundraising's New Era. Information Today, Inc. 2016. URL: https://www.infotoday.com/OnlineSearcher/Articles/Features/Crowdfunding-Research-and-Academic-Fundraisings-New-Era-113324.shtml?PageNum=1 [accessed 2020-09-23]

7. Bouncken RB, Komorek M, Kraus S. Crowdfunding: The Current State Of Research. IBER 2015 Apr 30;14(3):407. [doi: 10.19030/iber.v14i3.9206]

8. Hossain M, Oparaocha GO. Crowdfunding: Motives, Definitions, Typology and Ethical Challenges. Entrepreneurship Research Journal 2017 Jan 1;7(2):1-14. [doi: 10.1515/erj-2015-0045]

9. Bassani G, Marinelli N, Vismara S. Crowdfunding in healthcare. J Technol Transf 2018 Apr 20;44(4):1290-1310. [doi: 10.1007/s10961-018-9663-7]

10. Wheat RE, Wang Y, Byrnes JE, Ranganathan J. Raising money for scientific research through crowdfunding. Trends Ecol Evol 2013 Feb;28(2):71-72. [doi: 10.1016/j.tree.2012.11.001] [Medline: 23219380]

11. Krittanawong C, Zhang HJ, Aydar M, Wang Z, Sun T. Crowdfunding for cardiovascular research. Int J Cardiol 2018 Jan 01;250:268-269. [doi: 10.1016/j.ijcard.2017.10.029] [Medline: 29074042]

12. Kraus S, Richter C, Brem A, Cheng C, Chang M. Strategies for reward-based crowdfunding campaigns. 2016 Jan;1(1):13-23. [doi: 10.1016/j.jik.2016.01.010]

13. Belleflamme P, Omrani N, Peitz M. The economics of crowdfunding platforms. Information Economics and Policy 2015 Dec;33:11-28. [doi: 10.1016/j.infoecopol.2015.08.003]

14. Agrawal A, Catalini C, Goldfarb A. Some Simple Economics of Crowdfunding. Innovation Policy and the Economy 2014 Jan;14:63-97. [doi: 10.1086/674021]

15. Stummer W, Pichlmeier U, Meinel T, Wiestler OD, Zanella F, Reulen H. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. The Lancet Oncology 2006 May;7(5):392-401. [doi: 10.1016/S1470-2045(06)70665-9] [Medline: 16648043]

16. wemakeit. URL: https://wemakeit.com [accessed 2020-09-23]

17. Berendt M, Farquhar RG, Mandigers PJ, Pakozdy A, Bhatti SF, De Risio L, et al. International veterinary epilepsy task force consensus report on epilepsy definition, classification and terminology in companion animals. BMC Vet Res 2015 Aug 28;11:182 [FREE Full text] [doi: 10.1186/s12917-015-0461-2] [Medline: 26316133]

18. Dagal A, Lam AM. Cerebral autoregulation and anesthesia. Curr Opin Anaesthesiol 2009 Oct;22(5):547-552. [doi: 10.1097/ACO.0b013e32833020be] [Medline: 19620861]

19. Aleksina A, Akulenka S, Lublóy Á. Success factors of crowdfunding campaigns in medical research: perceptions and reality. Drug Discov Today 2019 Jul;24(7):1413-1420. [doi: 10.1016/j.drudis.2019.05.012] [Medline: 31132416]

20. Chen S, Thomas S, Kohli C. What Really Makes a Promotional Campaign Succeed on a Crowdfunding Platform? JAR 2016 Feb 12;56(1):81-94. [doi: 10.2501/jar-2016-002]

21. Koole MA, Kauw D, Winter MM, Schuuring MJ. A successful crowdfunding project for eHealth research on grown-up congenital heart disease patients. Int J Cardiol 2018 Dec 15;273:96-99 [FREE Full text] [doi: 10.1016/j.ijcard.2018.10.015] [Medline: 30297187]

22. Mollick E. The dynamics of crowdfunding: An exploratory study. Journal of Business Venturing 2014 Jan;29(1):1-16. [doi: 10.1016/j.jbusvent.2013.06.005]

XSL•FO

RenderX

23.   Kuppuswamy V, Bayus BL. Does my contribution to your crowdfunding project matter? Journal of Business Venturing
       2017 Jan;32(1):72-89. [doi: 10.1016/j.jbusvent.2016.10.004]
24.   Hong Y, Hu Y, Burtch G. Embeddedness, prosociality, and social influence: evidence from online crowdfunding. MIS
       Quarterly 2018. [doi: 10.25300/MISQ/2018/14105]
25.   Zhang H, Chen W. Crowdfunding technological innovations: Interaction between consumer benefits and rewards.
       Technovation 2019 Jun;84-85:11-20. [doi: 10.1016/j.technovation.2018.05.001]
26.   Sauermann H, Franzoni C, Shafi K. Crowdfunding scientific research: Descriptive insights and correlates of funding success.
       PLoS One 2019;14(1):e0208384 [FREE Full text] [doi: 10.1371/journal.pone.0208384] [Medline: 30608943]
27.   Xu B, Zheng H, Xu Y, Wang T. Configurational paths to sponsor satisfaction in crowdfunding. Journal of Business Research
       2016 Feb;69(2):915-927. [doi: 10.1016/j.jbusres.2015.06.040]

Abbreviations

CHF: Swiss francs

XSL•FO

RenderX

Original Paper

Potential of HIV Self-Sampling to Increase Testing Frequency Among Gay, Bisexual, and Other Men Who Have Sex With Men, and the Role of Online Result Communication: Online Cross-Sectional Study

Tomás Maté[1], MD, PhD; Juan Hoyos[2,3], PhD; Juan Miguel Guerras[2,4], BSc, MPH; Cristina Agustí[2,5], BSc, PhD; Sophocles Chanos[6], BA, BSc, MSc; Matthias Kuske[7], MA; Ricardo Fuertes[8], BSc; Roxana Stefanescu[9], BSc; Jose Pulido[2,3], PhD; Luis Sordo[2,3], MD, MPH, PhD; Luis de la Fuente[2,4], MD, MPH, PhD; María José Belza[2,10], MPH, PhD; EURO HIV EDAT Group[11]

[1]Gerencia Regional de Salud de Castilla y León, Valladolid, Spain

[2]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

[3]Departamento de Salud Pública y Materno-Infantil, Universidad Complutense de Madrid, Madrid, Spain

[4]Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Madrid, Spain

[5]Departament de Salut, Centre Estudis Epidemiologics sobre les Infeccions de Transmissio Sexual i Sida de Catalunya (CEEISCAT), Generalitat de Catalunya, Badalona, Spain

[6]Checkpoint Athens, Athens, Greece

[7]AIDS Hilfe NRW e.V., Berlin, Germany

[8]GAT-Grupo de Ativistas em Tratamentos, Lisboa, Portugal

[9]ARAS-Asociatia Romana Anti-SIDA, Bucharest, Romania

[10]Escuela Nacional de Sanidad, Instituto de Salud Carlos III, Madrid, Spain

[11]See Acknowledgments

Corresponding Author:
Juan Hoyos, PhD
CIBER Epidemiología y Salud Pública (CIBERESP)
Escuela Nacional de Sanidad. Pabellón 8, Instituto de Salud Carlos III
C/ Monforte de Lemos 5
Madrid, 28029
Spain
Phone: 34 91 8222 056
Email: hoyosmiller@hotmail.com

Abstract

Background: Late HIV diagnosis remains frequent among the gay, bisexual, and other men who have sex with men (GBMSM) population across Europe. HIV self-sampling could help remove barriers and facilitate access to testing for this high-risk population.

Objective: We assessed the capacity of HIV self-sampling to increase the testing frequency among GBMSM living in Denmark, Germany, Greece, Portugal, Romania, and Spain, and evaluated the role of new technologies in the result communication phase.

Methods: We analyzed a convenience sample of 5019 GBMSM with prior HIV testing experience who were recruited during 2016 through gay dating websites. We estimated the proportion of GBMSM who reported that the availability of self-sampling would result in an increase of their current testing frequency. We constructed a Poisson regression model for each country to calculate prevalence ratios and 95% CIs of factors associated with an increase of testing frequency as a result of self-sampling availability.

Results: Overall, 59% (between country range 54.2%-77.2%) of the participants considered that they would test more frequently for HIV if self-sampling was available in their country. In the multivariate analysis, the increase of testing frequency as a result of self-sampling availability was independently associated with reporting a higher number of unprotected anal intercourse events in all countries except for Greece. Independent associations were also observed among GBMSM who were not open about their sex life in Germany, Greece, Portugal, and Spain; those with a lower number of previous HIV tests in Denmark, Greece, Portugal,

and Spain; and for those that took their last test more than 3 months previously in Germany, Portugal, Romania, and Spain. In addition, 58.4% (range 40.5%-73.6%) of the participants indicated a preference for learning their result through one-way interaction methods, mainly via email (25.6%, range 16.8%-35.2%) and through a secure website (20.3%, range 7.3%-23.7%). Almost two thirds (65%) of GBMSM indicated preferring one of these methods even if the result was reactive.

Conclusions: Availability of HIV self-sampling kits as an additional testing methodology would lead to a much-needed increase of testing frequency, especially for the hidden, high-risk, and undertested GBMSM population. Online-based technologies without any personal interaction were preferred for the communication of the results, even for reactive results.

Introduction

HIV remains a significant public health problem in Europe. In 2017, 26,614 individuals were diagnosed as having HIV in the European Union/European Economic Area. Sex between men remains the most common transmission route and accounts for half of all new diagnoses [1]. A keystone of international strategies designed to curtail the HIV epidemic is to expand testing to reduce the undiagnosed fraction of the epidemic [2,3]. Estimates in 2018 indicated that there were 776,086 and 31,427 people living with HIV in western and central Europe, 13% and 17% of whom remained undiagnosed, respectively [4]. Delayed diagnosis remains common across Europe, and 53% of all new diagnoses are made at a late stage of infection (CD4 count of <350 mm^3). Although gay, bisexual, and other men who have sex with men (GBMSM) are less affected by a delayed diagnosis in relative terms (41.0%), they comprise the group with the highest absolute number of delayed diagnoses [1].

The promotion of HIV testing is in fact a vital point of the 90-90-90 UNAIDS plan to end the AIDS epidemic. According to this plan, by 2030, 90% of all people living with HIV should have a diagnosis, 90% of the diagnosed population should be on treatment, and 90% of these should achieve viral suppression [2]. The benefits of promoting earlier diagnosis are two-fold. From a public health perspective, testing enables early access to highly active antiretroviral therapy that leads to the reduction of viral load and onward transmission [5]. Receiving an HIV diagnosis also leads to short-term behavioral changes, and individuals tend to reduce their engagement in sexual risk behaviors [6]. From the patient perspective, early diagnosis has benefits of reducing morbidity and mortality [7,8]. In fact, if HIV is diagnosed and treated early, the life expectancy is similar to that of the general population [8].

In light of the importance of HIV testing, several strategies have been implemented in the last decade both in clinical and nonclinical settings [9,10]. In the case of GBMSM, recommendations are that they should be tested at least every 12 months [11]; however, a large proportion of this population still does not meet the proposed testing frequency in Europe [12]. Thus, there is still room for improvement. One of the strategies that has been rolled out to promote earlier diagnosis is HIV self-sampling, also known as home sampling or postal sampling. In this testing methodology, an individual collects his/her own blood or saliva sample using a suitable kit. The sample will then be posted to a designated laboratory for processing. Results are then delivered by phone, text message, online, or face to face.

With the exception of examples in the United Kingdom [13], Belgium [14,15], and Spain [16], this strategy has rarely been studied in Europe [17], and its capacity for increasing the HIV testing frequency among GBMSM remains unknown. Additionally, a key element that needs to be factored in when considering the introduction of self-sampling as an additional testing strategy is the preference regarding methods of result communication and if it would change depending on the result received. This aspect has never been studied and would help to shape the design of this testing strategy to the different national scenarios so as to achieve optimal linkage to HIV care.

In this study, we analyzed how self-sampling would contribute to the increase of testing frequency among GBMSM who were recruited online in 6 European countries where this testing option is yet to be introduced at a national scale. Additionally, we assessed the preferred methods of communicating test results taking into account the reactiveness of the self-tested sample.

Methods

Design

In the context of the EURO HIV EDAT project (Operational Knowledge to Improve HIV Early Diagnosis and Treatment Among Vulnerable Groups in Europe; Grant Agreement number 2013 11 01), we conducted an online cross-sectional study in 8 European countries (Belgium, Denmark, Greece, Germany, Portugal, Romania, Slovenia, and Spain) between April and December 2016. The working team included members of academia and local organizations that ran community-based voluntary counseling and testing sites. Further information on this project can be found elsewhere [18]. The project received approval of the Ethical Committee of Investigation and Animal Welfare of the Instituto de Salud Carlos III (CEI PI 52_2015-v2) and Hospital Germans Trias i Pujol (PI-14-106).

Recruitment Procedures

We aimed at recruiting a large convenience sample of GBMSM. The dissemination of an open survey was largely performed through GBMSM geospatial dating apps and websites that were previously identified by the local community-based voluntary

counseling and testing organizations in each participating country. Likewise, we made an effort to disseminate the survey in websites of GBMSM-oriented associations and supportive organizations. Advertisement was performed through banners, direct messages, and mailing lists. Those who clicked on the promotional banner or link were directed to an introductory screen that included brief information on the aims of the project, its anonymity, funding, and the partners involved. It also included a link to the EURO HIV EDAT project website for those interested in more information. To ensure complete anonymity, no cookies or internet protocols were collected. Participation was voluntary and those who decided to go ahead gave their informed consent by checking a box with the message "I have read and understood the above information, in the country I live in I am old enough to legally have sex, and I want to participate" before moving on to the first question. No incentives were offered for participation.

Inclusion and Exclusion Criteria

A total of 5799 HIV-negative GBMSM who resided in one of the participating countries accessed the questionnaire. We excluded 496 participants who had missing data on the question assessing our main outcome (potential of self-sampling to increase current testing frequency). Thus, the response rate was 91.45% (5303/5799). Additionally, we had to exclude participants from Slovenia (n=175) and Belgium (n=109) since the chosen dissemination websites in these two countries were not appropriately efficient, and the sample size was insufficient to run the multivariate analysis. Our final sample therefore comprised 5019 GBMSM from 6 countries.

Data Collection Instrument

Data were collected through a self-administered online questionnaire. The questionnaire was designed in English and translated to the language of each of the analyzed countries: Danish, German, Greek, Portuguese, Romanian, and Spanish. The translation was carried out by the partners of each participating country, all of whom were native speakers. Upon completion, questionnaires were back-translated using Google Translate to check for mistakes and inconsistencies with the English version.

The questionnaire included 90 items (1 item per page) although not every participant had to answer all questions. The completion time was approximately 20 minutes, which included sections to assess sociodemography, HIV testing history, outness about sex life with other men, sexual risk behaviors, and past diagnosis of sexually transmitted infections. To assess outness, we asked the participants "How would you describe the way you live your sex life with men?" This question had four response options: "openly," "discreetly," "hidden," "in total secrecy." The questionnaire also included a section that assessed several aspects related to HIV self-sampling. Before starting this section, the following definition was presented to participants: "Regarding self-sampling, sometimes referred to as 'Postal Tests,' you use a 'kit' to take either a blood or a saliva sample. This will then be posted to a laboratory for analysis and the results will be fed back to you." Our main outcome was the potential of self-sampling to increase current testing frequency, which was assessed through the following question: "If HIV

self-sampling was to be made available, you would test…": (1) more times/more frequently than now, (2) about the same, or (3) less than now. This question was answered only by HIV-negative GBMSM.

We also assessed the role of new technologies in the delivery of test results using the following question: "When the laboratory has your results, how would you like to receive them?" Participants could choose one of the following answers: (1) by email, (2) through a secure website, (3) By SMS text messaging, (4) through a face-to-face consultation at a medical office, (5) through a face-to-face consultation at a community-based organization or a nongovernmental organization, (6) by phone call, and (7) other. Those who chose options 1 to 3 were also asked: "What if the result was reactive, how would you prefer to receive it then?" They then had to choose between the following options: (1) I would choose the same option, (2) I would rather receive a phone call, (3) I would rather attend a face-to-face consultation at a medical office, (4) I would rather attend a face-to-face consultation at a community-based organization, and (5) other. Respondents were able to review and change their answers through use of a "Back" button. The contents of the instrument can be found in Multimedia Appendix 1. Data collection was anonymous (no questions including personal identification were included) and confidential. Before the launch, the questionnaire was piloted and revised by partners from all countries. No randomization of questionnaire items was performed.

Data Analysis

We performed a descriptive analysis of the main characteristics of the sample by country of residence. The data for the total number of participants were weighted to adjust for the disproportionate distribution of the sample by country of recruitment. Weighting coefficients were calculated using the male population between 18 and 65 years old living in the participating countries in 2016. Data were extracted from EUROSTAT [17]. Only weighted percentages are presented for the weighted total population.

For each country, we calculated the prevalence of participants who reported that the availability of HIV self-sampling kits would result in an increase of their testing frequency by relevant sociodemographic, behavioral, and HIV testing history variables. To estimate the factors associated with an increase of the testing frequency due to HIV self-sampling, we conducted a Poisson regression with robust variance multivariate analysis to estimate the crude and adjusted prevalence ratios (PRs) with their 95% CIs. A multivariate model was built for each country. We chose Poisson regression with robust variance because it is a better alternative than logistic regression when working with frequent outcomes [19,20]. We initially included all of the relevant variables with a significance level ≤.20 and used the minimum Akaike information criterion and the minimum Bayesian Schwartz information criterion for model comparisons and to select the optimal model.

Finally, we assessed the preferred methods of result communication by country. We grouped methods in two groups depending on the information flow they allowed: (1) unidirectional methods, in which communication flows only

toward the user with no possibility of asking questions or giving direct feedback (email, secure website, SMS text messaging); and (2) bidirectional methods, in which communication can flow two ways (medical office, community-based organizations/nongovernmental organizations, phone call).

For those who chose a unidirectional method, we estimated the proportion that would have chosen the same testing method even if the result was reactive. Data on the total number of participants were also weighted using the aforementioned method.

Results

Participant Characteristics

The main demographic and HIV-related characteristics are summarized in Table 1. Among the total 5019 participants, 72.4% were between 25-49 years of age and 90.5% were born in their current country of residence. Approximately one third lived in cities with under 100,000 inhabitants, ranging from 26.4% (132/499) in Greece to 46.8% (247/528) in Portugal, and 47.3% had not finished a university degree at the time of the survey, ranging from 32.8% (165/503) in Greece to 56.5% (530/938) in Germany.

Overall, 20.2% of the participants indicated that they kept their sex life with other men hidden or in total secrecy, with the lowest proportion in Denmark and the highest in Romania. Regarding sexual risk indicators, 64.2% reported having had unprotected anal intercourse in the last 12 months with at least one partner, ranging from 48% (172/358) in Greece to 76.5% (186/243) in Denmark, and 41.8% had previously been diagnosed with a sexually transmitted infection, ranging from 31.2% (60/192) in Romania to 55.2% (134/243) in Denmark.

Overall, 19.9% self-reported having been tested for HIV at least once in the past, ranging from 13.1% (42/320) in Denmark to 28% (86/307) in Romania, and 39.1% had received their last test more than 12 months previously, ranging from 27% (136/502) in Greece to 43.1% (404/938) in Germany.

XSL•FO
RenderX

Table 1. Sociodemographic profile, outness, sexual behaviors, history of sexually transmitted infections (STIs), and testing history of the participants.

Characteristic	Denmark (N=320), n (%)	Germany (N=941), n (%)	Greece (N=503), n (%)	Portugal (N=531), n (%)	Romania (N=307), n (%)	Spain (N=2417), n (%)	Total[a] (N=5019)
Age (years)							
<25	35 (10.9)	57 (6.1)	79 (15.7)	46 (8.7)	44 (14.3)	255 (10.6)	9.1
25-29	48 (15.0)	108 (11.5)	78 (15.5)	66 (12.4)	59 (19.2)	395 (16.3)	14.1
30-34	41 (12.8)	132 (14.0)	81 (16.1)	84 (15.8)	64 (20.8)	388 (16.1)	15.5
35-39	54 (16.9)	149 (15.8)	111 (22.1)	95 (17.9)	45 (14.7)	364 (15.1)	16.0
40-49	93 (29.1)	276 (29.3)	104 (20.7)	139 (26.2)	62 (20.2)	636 (26.3)	26.8
≥50	49 (15.3)	219 (23.3)	50 (9.9)	101 (19.0)	33 (10.7)	379 (15.7)	18.5
Place of birth							
In country of current residence	275 (86.2)	844 (90.5)	475 (97.3)	463 (87.5)	299 (98.0)	2076 (86.8)	90.5
Europe	26 (8.2)	58 (6.2)	7 (1.4)	17 (3.2)	4 (1.3)	100 (4.2)	4.7
Other	18 (5.6)	31 (3.3)	6 (1.2)	49 (9.3)	2 (0.7)	215 (9.0)	4.8
Number of inhabitants in place of residence							
≥1,000,000	111 (34.8)	260 (27.7)	269 (53.9)	104 (19.7)	87 (28.5)	865 (35.9)	31.3
100,000-999.999	104 (32.6)	336 (35.8)	98 (19.6)	177 (33.5)	130 (42.6)	844 (35.0)	35.2
10,000-99,999	66 (20.7)	233 (24.8)	98 (19.6)	169 (32.0)	58 (19.0)	515 (21.4)	23.2
<10,000	38 (11.9)	110 (11.7)	34 (6.8)	78 (14.8)	30 (9.8)	186 (7.7)	10.3
Education							
Upper secondary education	140 (43.9)	309 (32.9)	72 (14.3)	170 (32.1)	84 (27.7)	632 (26.2)	29.7
Postsecondary/nontertiary education	11 (3.4)	221 (23.6)	93 (18.5)	30 (5.7)	35 (11.6)	331 (13.7)	17.6
University education	168 (52.7)	408 (43.5)	338 (67.2)	330 (62.3)	184 (60.7)	1450 (60.1)	52.6
Lives sex life with men…							
Openly	244 (76.3)	498 (53.0)	94 (18.7)	107 (20.2)	38 (12.4)	1110 (45.9)	43.3
Discreetly	56 (17.5)	267 (28.4)	239 (47.5)	285 (53.7)	158 (51.5)	977 (40.4)	36.4
Hidden/In total secrecy	20 (6.3)	175 (18.6)	170 (33.8)	139 (26.2)	111 (36.2)	329 (13.6)	20.2
Number of partners with unprotected anal intercourse (last 12 months)							
None	57 (23.5)	255 (36.2)	186 (52.0)	136 (33.4)	54 (27.6)	692 (37.1)	35.8
1	63 (25.9)	208 (29.5)	116 (32.4)	127 (31.2)	73 (37.2)	613 (32.8)	31.4
2-4	66 (27.2)	153 (21.7)	43 (12.0)	100 (24.6)	49 (25.0)	372 (19.9)	21.4
≥5	57 (23.5)	89 (12.6)	13 (3.6)	44 (10.8)	20 (10.2)	190 (10.2)	11.4
History of STI diagnosis							
Last 12 months	34 (14.0)	70 (10.0)	41 (11.6)	59 (14.8)	15 (7.8)	203 (11.0)	10.5
>12 months ago	100 (41.2)	229 (32.8)	90 (25.6)	111 (27.9)	45 (23.4)	608 (32.9)	31.3
None	109 (44.9)	399 (57.2)	221 (62.8)	228 (57.3)	132 (68.8)	1038 (56.1)	58.2
Number of HIV tests (ever)							
1	42 (13.1)	176 (18.7)	95 (18.9)	97 (18.3)	86 (28.0)	484 (20.0)	19.9
2-5	122 (38.1)	447 (47.5)	239 (47.5)	243 (45.8)	154 (50.2)	1144 (47.3)	47.4
6-9	64 (20.0)	148 (15.7)	71 (14.1)	87 (16.4)	31 (10.1)	382 (15.8)	15.2
≥10	92 (28.8)	170 (18.1)	98 (19.5)	104 (19.6)	36 (11.7)	407 (16.8)	17.5

XSL•FO
RenderX

Characteristic	Denmark (N=320), n (%)	Germany (N=941), n (%)	Greece (N=503), n (%)	Portugal (N=531), n (%)	Romania (N=307), n (%)	Spain (N=2417), n (%)	Total[a] (N=5019)
Time since last HIV test							
≤3 months	84 (26.3)	197 (21.0)	184 (36.7)	112 (21.1)	69 (22.5)	611 (25.4)	23.4
3-12 months	125 (39.2)	337 (35.9)	182 (36.3)	238 (44.9)	117 (38.1)	921 (38.2)	37.4
1-2 years	57 (17.9)	165 (17.6)	69 (13.7)	87 (16.4)	60 (19.5)	433 (18.0)	17.6
>2 years	53 (16.6)	239 (25.5)	67 (13.3)	93 (17.5)	61 (19.9)	445 (18.5)	21.5

[a] Weighted percentages, calculated exclusively using valid data.

Factors Associated With Potential of Self-Sampling to Increase Current Testing Frequency

Regarding the potential of self-sampling for increasing the current frequency of HIV testing, 59% (54.2% in Spain, 55.9% in Germany, 59.1% in Denmark, 63.0% in Greece, 66.7% in Portugal, and 77.2% in Romania) indicated that, if available in their country of residence, they would test more frequently. Proportions of participants reporting an increase of testing frequency as a result of the availability of self-sampling by relevant variables can be found in Multimedia Appendix 2.

Table 2 shows the results of the multivariate analysis performed for each of the 6 countries separately. The variables that were independently associated with the potential of self-sampling to increase testing frequency were outness about sex life with other men, number of partners with whom they had unprotected anal intercourse in the last 12 months, number of lifetime HIV tests performed, and time since their last HIV test.

With the exception of GBMSM living in Denmark and Romania, the potential of self-sampling to increase testing frequency was higher among those who lived their sex life discreetly or "hidden/in total secrecy" than among participants who were open about their sex lives.

The number of partners with whom participants reported having engaged in unprotected anal intercourse was also associated with the capacity of self-sampling to increase testing frequency in all countries except for Greece. We found significant adjusted PRs among Spain-based respondents who reported having engaged in unprotected anal intercourse with only one partner. In all countries, the adjusted PRs were statistically significant among those who reported being involved in unprotected anal intercourse with 2-4 partners and the PRs increased further among those who reported having had ≥5 partners in the last 12 months. The adjusted PR decreased with the lifetime number of HIV tests. This pattern was the same in all countries except for Germany and Romania, in which no statistical significance was found.

Finally, apart from Denmark and Greece, the potential of self-sampling to increase testing frequency was independently associated with time elapsed since the last HIV test. Compared to those tested ≤3 months ago, the adjusted PR gradually increased among GBMSM who were last tested between 3 and 12 months prior and peaked among those tested >12 months previously.

Table 2. Factors associated with a potential increase of testing frequency if self-sampling was made available.

Variable	Denmark (N=320), PR[a] (95% CI)	Germany (N=941), PR (95% CI)	Greece (N=503), PR (95% CI)	Portugal (N=531), PR (95% CI)	Romania (N=307), PR (95% CI)	Spain (N=2417), PR (95% CI)
Age (years)						
≥50	1.0 (0.7-1.4)	1.0 (0.8-1.3)	1.2 (1.0-1.5)	1.1 (0.8-1.4)	1.0 (0.8-1.2)	1.1 (0.9-1.3)
25-49	1.0 (0.8-1.3)	1.1 (0.8-1.4)	1.1 (0.9-1.3)	1.3 (1.0-1.6)	1.0 (0.9-1.2)	1.2 (1.0-1.3)
<25	1.0	1.0	1.0	1.0	1.0	1.0
Education						
No university education	1.3 (1.1-1.5)	—[b]	—	—	—	—
University education	1.0	—	—	—	—	—
Lives sex life with men…						
Hidden/Total secrecy	—	1.4 (1.2-1.6)	1.4 (1.1-1.8)	1.2 (1.0-1.5)	—	1.3 (1.1-1.4)
Discreetly	—	1.3 (1.1-1.4)	1.2 (0.9-1.5)	1.0 (0.9-1.3)	—	1.2 (1.1-1.3)
Openly	—	1.0	1.0	1.0	—	1.0
Number of partners with unprotected anal intercourse (last 12 months)						
≥5	1.5 (1.1-2.1)	1.3 (1.0-1.6)	—	1.3 (1.1-1.7)	1.3 (1.1-1.6)	1.5 (1.3-1.7)
2-4	1.4 (1.0-2.0)	1.3 (1.1-1.5)	—	1.3 (1.1-1.5)	1.3 (1.1-1.6)	1.3 (1.2-1.5)
1	1.1 (0.8-1.6)	1.1 (0.9-1.3)	—	1.1 (0.9-1.4)	1.1 (0.9-1.3)	1.1 (1.0-1.3)
None	1.0	1.0	—	1.0	1.0	1.0
Number of HIV tests (ever)						
1	1.5 (1.2-2.0)	—	1.7 (1.3-2.2)	1.7 (1.3-2.1)	—	1.5 (1.3-1.8)
2-5	1.3 (1.0-1.7)	—	1.7 (1.3-2.1)	1.5 (1.2-2.0)	—	1.4 (1.2-1.6)
6-9	1.2 (0.9-1.6)	—	1.3 (0.9-1.8)	1.3 (1.0-1.8)	—	1.1 (1.0-1.3)
≥10	1.0	—	1.0	1.0	—	1.0
Time since last HIV test						
>12 months	—	1.3 (1.1-1.6)	—	1.4 (1.1-1.8)	1.4 (1.1-1.7)	1.3 (1.1-1.4)
3-12 months	—	1.1 (0.9-1.3)	—	1.3 (1.1-1.6)	1.4 (1.1-1.7)	1.1 (1.0-1.3)
≤3 months	—	1.0	—	1.0	1.0	1.0

[a]PR: adjusted prevalence ratio.

[b]Variable not included in multivariate analysis.

Preferences Regarding Method of Test Result Communication

Table 3 presents the results of the participants' preferences regarding the method of result communication. The majority of respondents preferred unidirectional methods of communication (email, secure website, and SMS text messaging). The preferred bidirectional method was face-to-face consultation at a medical office and the least preferred method was a phone call. Nearly two-thirds of those who chose unidirectional methods indicated that they would retain this preference over other methods even if the self-sampling result was determined to be reactive.

Table 3. Preferred HIV test result communication method in general and for a reactive result.

Communication Method	Denmark (N=320), n (%)	Germany (N=941), n (%)	Greece (N=503), n (%)	Portugal (N=531), n (%)	Romania (N=307), n (%)	Spain (N=2417), n (%)	Total[a] (N=5019)
Unidirectional							
All	151 (52.8)	467 (53.6)	195 (40.5)	377 (73.6)	187 (62.5)	1518 (66.4)	58.4
Email							
Overall	48 (16.8)	179 (20.6)	88 (18.3)	180 (35.2)	96 (32.1)	738 (32.3)	25.6
Same response if result is reactive	30 (62.5)	126 (70.4)	52 (59.1)	120 (66.7)	61 (63.5)	437 (59.2)	65.7
Secure website							
Overall	48 (16.8)	206 (23.7)	35 (7.3)	109 (21.3)	46 (15.4)	445 (19.5)	20.3
Same response if result is reactive	35 (72.9)	144 (69.9)	20 (57.1)	73 (67.0)	29 (63.0)	254 (57.1)	65.0
SMS text message							
Overall	55 (19.2)	82 (9.4)	72 (14.9)	88 (17.2)	45 (15.1)	335 (14.7)	12.5
Same response if test is reactive	33 (60.0)	53 (64.6)	38 (52.8)	57 (64.8)	30 (66.7)	186 (55.5)	62.1
Bidirectional							
All	132 (46.2)	371 (42.6)	282 (58.5)	131 (25.6)	105 (35.1)	729 (31.9)	39.0
Medical office	55 (19.2)	220 (25.3)	139 (28.8)	66 (12.9)	75 (25.1)	522 (22.8)	23.9
At a CBO[b]/NGO[c]	33 (11.5)	90 (10.3)	82 (17.0)	52 (10.2)	10 (3.3)	80 (3.5)	8.1
Phone call	44 (15.4)	61 (7.0)	61 (12.7)	13 (2.5)	20 (6.7)	127 (5.6)	6.9
Other	3 (1.0)	33 (3.8)	5 (1.0)	4 (0.8)	7 (2.3)	39 (1.7)	2.7

[a]Weighted percentages, calculated exclusively using valid data.

[b]CBO: community-based organization.

[c]NGO: nongovernmental organization.

Discussion

Our study shows that HIV self-sampling has high potential to increase the frequency of HIV testing among GBMSM in the 6 European countries studied. More than half of the respondents reported that, if made available, self-sampling would increase their testing frequency. This potential was found to be especially high among those who are not open about their sexuality, who had a greater number of sexual partners, and for those who are undertested. Based on the preferences expressed by the participants, online technologies should play a key role in communicating test results.

The evidence base on HIV self-sampling kits is still very weak. Two studies from Belgium [14,15] and one study from Spain [16] have proven that self-sampling kits can be used in outreach activities at locations frequented by populations at high risk, including GBMSM [14,16]. In the United Kingdom, the national HIV self-sampling service, which is fully online-based, has been running since 2015 [13]. Outside of these three countries, the only European study we have found that assessed this diagnostic strategy was also based on the EURO HIV EDAT project and concluded that self-sampling, although not very well known, had high potential of use in European countries [17]. This is the first study to address the question as to whether

the introduction of this method would result in an increase of testing frequency. Our data suggest that the availability of HIV-self sampling kits would lead to increased testing rates in several key subpopulations.

Subpopulations who do not live their sex life openly are hard to reach and are the least likely to encounter or access health promotion interventions [12]. Thus, GBMSM who are the least open about their sex life generally presented lower testing rates in the last 12 months than the rest of the GBMSM [12]. Some characteristics of self-sampling methodologies such as privacy and not having to reveal sexual orientation or discuss their sex life with a clinician could facilitate testing for GBMSM that have yet to come to terms with their own sexuality. Based on our data, this hypothesis is plausible since the reported potential of self-sampling kits to increase testing frequency was especially high among GBMSM who are not completely open about their sex lives.

The ultimate goal of the introduction of strategies such as self-sampling is to promote testing and facilitate testing to undertested populations. Current testing recommendations for GBMSM are demanding, and barriers such as fear of stigma and discrimination, lack of anonymity or confidentiality, and waiting times at testing sites [21] have been described as deterrents to testing. Our results suggest that self-sampling could

XSL•FO
RenderX

help to alleviate some of these barriers, leading to an increase of the number of tests performed and less time between testing in undertested individuals. This could mean that a fraction of those who are not meeting testing recommendations could do so if self-sampling kits are available, allowing for seroconversions to be detected earlier.

From a public health perspective, shortening the time from infection to diagnosis by increasing testing frequency is particularly relevant in populations with a high number of sexual partners. Those who reported being involved in unprotected anal intercourse with a higher number of partners presented higher probabilities of unknowingly transmitting or acquiring HIV from partners of a serodiscordant status. According to our results, the introduction of self-sampling would increase testing frequency precisely in this key population. In this study, we only considered condom use during anal intercourse and we did not assess other risk reduction strategies such as serosorting. Those reporting a higher number of unprotected anal intercourse events could be having unprotected anal sex only with partners of a presumably equal (negative) serostatus (serosorting) or with HIV-positive partners with a presumably undetectable viral load to avoid infection [22,23]. Nevertheless, the effectiveness of serosorting is questionable [24-26] and the percentage of virally suppressed HIV-positive GBMSM varies widely across countries [4]; thus, the impact of self-sampling in the testing frequency of GBMSM who use this protective strategy needs to be better understood.

Additionally, we did not assess HIV pre-exposure prophylaxis (PREP). Those involved in unprotected anal intercourse could be receiving PREP but this is unlikely since it was not widely available at the time of data collection (2016). Nevertheless, PREP has now been incorporated into the preventive options among GBMSM, and the role that HIV self-sampling could play in the follow-up of patients on PREP merits further study. Conversely, postexposure prophylaxis is an available option and is an effective means of protection against HIV. We also did not assess if some of those involved in unprotected sex could be doing so with the intention of taking postexposure prophylaxis after the risk exposure; however, this is highly unlikely since uptake is very low among GBMSM across Europe [12].

When assessing the introduction of self-sampling, it is important to consider several steps along the testing process, namely the distribution of the testing kits and the consultation of the test result. The aforementioned examples of studies assessing self-sampling kits distributed during outreach activities in locations frequented by GBMSM, such as saunas or gay venues, were able to conduct a modest number of tests [14-16]. Consequently, the number of new HIV diagnoses was also limited. Another approach is to embed the distribution of self-sampling kits in already existing health services and community organizations. There is a gap of knowledge regarding the acceptability of this approach for GBMSM, but an assessment focusing on black Africans of all sexualities concluded that, for several reasons, the number of self-sampling kits distributed and conducted based on this distribution pathway was also very low [27]. In this sense, methodologies fully based on internet distribution have proven to be by far the most efficient. In the United Kingdom, this methodology has facilitated testing for more than 69,000 self-samples, 788 of which were found to be reactive [13]. When introducing HIV self-sampling in national testing policies, countries need to first consider online distribution. This is particularly essential in the case of GBMSM considering that online gay dating apps and websites offer an extremely efficient platform to reach out to individuals for conducting public health interventions.

Timely linkage to care following an HIV diagnosis is also critical since late access can result in worse patient outcomes [28]. This is particularly relevant when considering testing methodologies outside of clinical settings such as that assessed in this study; thus, countries need to establish robust confirmation routes to assure optimal linkage to care. In self-sampling methodologies, result consultation needs to be taken into account. With the exception of Greece, internet-based methods were preferred by our participants, even in the case of receiving a reactive result. These preferences need to be considered since it is of utter importance that tested individuals obtain further consultation of their results, especially if reactive, as this is the first step for result confirmation and consequent linkage to care. The use of digital communication technologies has proven to be effective in promoting testing among GBMSM [29]; however, studies are needed to assess if they are also capable of providing timely confirmation and linkage to care for those using self-sampling methodologies.

Our results are not without limitations. Our sample was mainly recruited via gay dating apps and websites. Although these platforms are increasingly used by GBMSM as a way of socializing and meeting new sexual partners, generalization of these results to the overall GBMSM population needs to be made with caution, especially given the limited sample size in some countries. In this sense, GBMSM identifying as gay and reporting more sexual risk behaviors could be overrepresented in our sample as has been previously reported [30]. Another limitation of this study is that we are working on a hypothetical situation. Although there is some evidence suggesting that answers based on hypothetical situations are able to predict actual behaviors [31-33], it is unclear if this is also the case with respect to the capacity of self-sampling to increase testing frequency. The definition of self-sampling included in the questionnaire explicitly mentioned the two types of self-sampling kits: saliva and blood-based kits. Thus, participants likely gave their answer considering both samples as potential options. This should also be taken into account when interpreting our results. For example, it is possible that if only one of the two sample options was made available, the proportion of participants reporting an increase of testing frequency could have been lower. Price (if marketed) and implementation problems resulting in invalid self-samples or delays in the communication of results could affect the capacity of HIV self-sampling to increase testing frequency. These barriers have not been assessed and merit further study.

As a way of ensuring complete anonymity and confidentiality, we did not collect internet protocols or cookies, and therefore could not assess the possibility of an individual answering the questionnaire more than once. However, the overall objective of the survey was clearly explained in the access screen, and

given that no compensation was provided in exchange for participation, this situation was highly unlikely. An important strength is that the assessment of HIV self-sampling was performed in a set of countries that had not introduced this testing option at the time of conducting the survey. This provides baseline information to policymakers that could guide them in the introduction of this methodology.

Based on our results, national HIV testing policies in the 6 countries evaluated in this study should consider the

incorporation of HIV self-sampling as an additional testing option, since its introduction could increase the testing frequency in a high proportion of GBMSM, especially among those who are not open about their sex lives, who remain undertested, and who report a high frequency of high-risk sexual behaviors. When designing its implementation, priority needs to be given to online methods of result communication, accompanied by clear referral pathways to HIV care for those obtaining a reactive result.

Conflicts of Interest

CA has received financial support for research and travel reimbursements from Gilead Sciences and ViiV Healthcare. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 162 KB - jmir_v22i11e21268_app1.docx]

Multimedia Appendix 2
Proportion of participants who reported that HIV self-sampling would increase their testing frequency by relevant variables.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v22i11e21268_app2.xlsx]

References

1. HIV/AIDS surveillance in Europe 2019. 2018 data. European Centre for Disease Prevention and Control and World Health Organization Regional Office for Europe. 2019. URL: https://www.ecdc.europa.eu/sites/default/files/documents/hiv-surveillance-report-2019.pdf [accessed 2020-11-08]
2. Joint United Nations Programme on HIV/AIDS. 90-90-90. An ambitious target to help end the AIDS epidemic. UNAIDS. 2014. URL: https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf [accessed 2020-11-08]
3. European Centre for Disease Prevention and Control. HIV and STI prevention among men who have sex with men. ECDC Guidance. 2015. URL: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/hiv-sti-prevention-among-men-who-have-sex-with-men-guidance.pdf [accessed 2020-11-08]
4. Special Report: Continuum of HIV care. Monitoring implementation of the Dublin Declaration on partnership to fight HIV/AIDS in Europe and Central Asia: 2018 progress report. European Centre for Disease Prevention and Control. 2018. URL: https://www.ecdc.europa.eu/sites/default/files/documents/HIV-continuum-of-care-monitoring-dublin-declaration-progress-report-2018.pdf [accessed 2020-11-08]
5. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
6. Marks G, Crepaz N, Senterfitt JW, Janssen RS. Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. J Acquir Immune Defic Syndr 2005 Aug 01;39(4):446-453. [doi: 10.1097/01.qai.0000151079.33935.79] [Medline: 16010168]

XSL•FO
RenderX

7.   Kitahata MM, Gange SJ, Abraham AG, Merriman B, Saag MS, Justice AC, NA-ACCORD Investigators. Effect of early versus deferred antiretroviral therapy for HIV on survival. N Engl J Med 2009 Apr 30;360(18):1815-1826 [FREE Full text] [doi: 10.1056/NEJMoa0807252] [Medline: 19339714]

8.   Rodger AJ, Lodwick R, Schechter M, Deeks S, Amin J, Gilson R, INSIGHT SMART, ESPRIT Study Groups. Mortality in well controlled HIV in the continuous antiretroviral therapy arms of the SMART and ESPRIT trials compared with the general population. AIDS 2013 Mar 27;27(6):973-979. [doi: 10.1097/QAD.0b013e32835cae9c] [Medline: 23698063]

9.   Desai S, Tavoschi L, Sullivan AK, Combs L, Raben D, Delpech V, et al. HIV testing strategies employed in health care settings in the European Union/European Economic Area (EU/EEA): evidence from a systematic review. HIV Med 2020 Mar;21(3):163-179 [FREE Full text] [doi: 10.1111/hiv.12809] [Medline: 31729150]

10.   Croxford S, Tavoschi L, Sullivan AK, Combs L, Raben D, Delpech V, et al. HIV testing strategies outside of health care settings in the European Union (EU)/European Economic Area (EEA): a systematic review to inform European Centre for Disease Prevention and Control guidance. HIV Med 2020 Mar;21(3):142-162 [FREE Full text] [doi: 10.1111/hiv.12807] [Medline: 31682060]

11.   Scientific Advice: Public health guidance on HIV, hepatitis B and C testing in the EU/EEA. European Centre for Disease Prevention and Control. 2018. URL: https://www.ecdc.europa.eu/sites/default/files/documents/hiv-hep-testing-guidance_0.pdf [accessed 2020-11-08]

12.   Technical Report: The European Men-Who-Have-Sex-With-Men Internet Survey. Key findings from 50 countries. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/sites/default/files/documents/European-MSM-internet-survey-2017-findings.pdf [accessed 2020-11-08]

13.   National HIV self-sampling service. November 2017 to October. Public Health England. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/789793/national_HIV_self_sampling_service_november_2017_to_october_2018.pdf [accessed 2020-11-08]

14.   Platteau T, Fransen K, Apers L, Kenyon C, Albers L, Vermoesen T, et al. Swab2know: An HIV-Testing Strategy Using Oral Fluid Samples and Online Communication of Test Results for Men Who Have Sex With Men in Belgium. J Med Internet Res 2015 Sep 01;17(9):e213 [FREE Full text] [doi: 10.2196/jmir.4384] [Medline: 26330138]

15.   Loos J, Manirankunda L, Platteau T, Albers L, Fransen K, Vermoesen T, et al. Acceptability of a Community-Based Outreach HIV-Testing Intervention Using Oral Fluid Collection Devices and Web-Based HIV Test Result Collection Among Sub-Saharan African Migrants: A Mixed-Method Study. JMIR Public Health Surveill 2016 Aug 04;2(2):e33 [FREE Full text] [doi: 10.2196/publichealth.5519] [Medline: 27493067]

16.   Agustí C, Muñoz R, González V, Villegas L, Fibla J, Meroño M, et al. Outreach HIV testing using oral fluid and online consultation of the results: Pilot intervention in Catalonia. Enferm Infecc Microbiol Clin 2020 Mar 06:Online ahead of print. [doi: 10.1016/j.eimc.2020.01.020] [Medline: 32151468]

17.   Hoyos J, Maté T, Indave BI, Agustí C, Chanos S, Pichon F, et al. Knowledge, actual and potential use of HIV self-sampling testing kits among MSM recruited in eight European countries. HIV Med 2018 Feb;19(Suppl 1):27-33. [doi: 10.1111/hiv.12593] [Medline: 29488699]

18.   Hoyos J, Belza M, Fernández-Balbuena S, Pérez S, de la Fuente L, Arranz B, et al. KAB/P study on the implementation of innovative HIV testing strategies: Main results of a study conducted among MSM and stakeholders. EUROHIVEDAT project. 2017. URL: https://eurohivedat.eu/arxius/ehe_docsmenu_docsmenu_doc_154-Final_Report_KABP_study_on_the_implementation_of_innovative_HIV_testing_strategies.pdf [accessed 2020-11-08]

19.   Barros AJD, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med Res Methodol 2003 Oct 20;3:21 [FREE Full text] [doi: 10.1186/1471-2288-3-21] [Medline: 14567763]

20.   Lee J, Tan CS, Chia KS. A practical guide for multivariate analysis of dichotomous outcomes. Ann Acad Med Singap 2009 Aug;38(8):714-719 [FREE Full text] [Medline: 19736577]

21.   ECDC Guidance. HIV testing: increasing uptake and effectiveness in the European Union. European Centre for Disease Prevention and Control. 2010. URL: https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/101129_GUI_HIV_testing.pdf [accessed 2020-11-08]

22.   Eisinger RW, Dieffenbach CW, Fauci AS. HIV Viral Load and Transmissibility of HIV Infection: Undetectable Equals Untransmittable. JAMA 2019 Feb 05;321(5):451-452. [doi: 10.1001/jama.2018.21167] [Medline: 30629090]

23.   Rodger AJ, Cambiano V, Bruun T, Vernazza P, Collins S, Degen O, PARTNER Study Group. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. Lancet 2019 Jun 15;393(10189):2428-2438 [FREE Full text] [doi: 10.1016/S0140-6736(19)30418-0] [Medline: 31056293]

24.   Dosekun O, Fox J. An overview of the relative risks of different sexual behaviours on HIV transmission. Curr Opin HIV AIDS 2010 Jul;5(4):291-297. [doi: 10.1097/COH.0b013e32833a88a3] [Medline: 20543603]

25.   van den Boom W, Konings R, Davidovich U, Sandfort T, Prins M, Stolte IG. Is serosorting effective in reducing the risk of HIV infection among men who have sex with men with casual sex partners? J Acquir Immune Defic Syndr 2014 Mar 01;65(3):375-379 [FREE Full text] [doi: 10.1097/QAI.0000000000000051] [Medline: 24189150]

XSL•FO
RenderX

26. Khosropour CM, Dombrowski JC, Swanson F, Kerani RP, Katz DA, Barbee LA, et al. Trends in Serosorting and the Association With HIV/STI Risk Over Time Among Men Who Have Sex With Men. J Acquir Immune Defic Syndr 2016 Jun 01;72(2):189-197 [FREE Full text] [doi: 10.1097/QAI.0000000000000947] [Medline: 26885806]

27. Seguin M, Dodds C, Mugweni E, McDaid L, Flowers P, Wayal S, et al. Self-sampling kits to increase HIV testing among black Africans in the UK: the HAUS mixed-methods study. Health Technol Assess 2018 Apr;22(22):1-158. [doi: 10.3310/hta22220] [Medline: 29717978]

28. Croxford S, Yin Z, Burns F, Copas A, Town K, Desai S, OptTEST project. Linkage to HIV care following diagnosis in the WHO European Region: A systematic review and meta-analysis, 2006-2017. PLoS One 2018;13(2):e0192403 [FREE Full text] [doi: 10.1371/journal.pone.0192403] [Medline: 29451875]

29. Veronese V, Ryan KE, Hughes C, Lim MS, Pedrana A, Stoové M. Using Digital Communication Technology to Increase HIV Testing Among Men Who Have Sex With Men and Transgender Women: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 28;22(7):e14230 [FREE Full text] [doi: 10.2196/14230] [Medline: 32720902]

30. Prah P, Hickson F, Bonell C, McDaid LM, Johnson AM, Wayal S, et al. Men who have sex with men in Great Britain: comparing methods and estimates from probability and convenience sample surveys. Sex Transm Infect 2016 Sep;92(6):455-463 [FREE Full text] [doi: 10.1136/sextrans-2015-052389] [Medline: 26965869]

31. Lancsar E, Swait J. Reconceptualising the external validity of discrete choice experiments. Pharmacoeconomics 2014 Oct;32(10):951-965. [doi: 10.1007/s40273-014-0181-7] [Medline: 24920196]

32. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ 2018 Nov;19(8):1053-1066. [doi: 10.1007/s10198-018-0954-6] [Medline: 29380229]

33. Salampessy BH, Veldwijk J, Jantine Schuit A, van den Brekel-Dijkstra K, Neslo REJ, Ardine de Wit G, et al. The Predictive Value of Discrete Choice Experiments in Public Health: An Exploratory Application. Patient 2015 Dec;8(6):521-529 [FREE Full text] [doi: 10.1007/s40271-015-0115-2] [Medline: 25618790]

Abbreviations

GBMSM: gay, bisexual, and other men who have sex with men
PR: prevalence ratio
PREP: pre-exposure prophylaxis

XSL•FO

RenderX

<u>Original Paper</u>

Outcomes of Equity-Oriented, Web-Based Parenting Information in Mothers of Low Socioeconomic Status Compared to Other Mothers: Participatory Mixed Methods Study

Pierre Pluye[1], PhD, MD; Reem El Sherif[1], MSc, PhDc; Araceli Gonzalez-Reyes[1], MSc, PhDc; Emmanuelle Turcotte[1], PhD; Tibor Schuster[1], PhD; Gillian Bartlett[1], PhD; Roland M Grad[1], MDCM, MSc; Vera Granikov[2], MLiS, PhDc; Melanie Barwick[3], PhD, C Psych; Geneviève Doray[4], BA; François Lagarde[4], MA; Christine Loignon[5], PhD

[1]Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]School of Information Studies, McGill University, Montreal, QC, Canada

[3]Child Health Evaluative Sciences, Research Institute, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

[4]Lucie and André Chagnon Foundation, Montreal, QC, Canada

[5]Department of Family Medicine, Université de Sherbrooke, Sherbrooke, QC, Canada

Corresponding Author:
Pierre Pluye, PhD, MD
Department of Family Medicine
McGill University
5858 Cote des neiges, Rm 300
Montreal, QC, H3S 1Z1
Canada
Phone: 1 514 398 8483
Email: pierre.pluye@mcgill.ca

Abstract

Background: Typically, web-based consumer health information is considered more beneficial for people with high levels of education and income. No evidence shows that equity-oriented information offers equal benefits to all. This is important for parents of low socioeconomic status (SES; low levels of education and income and usually a low level of literacy).

Objective: This study is based on a conceptual framework of information outcomes. In light of this, it aims to compare the perception of the outcomes of web-based parenting information in low-SES mothers with that of other mothers and explore the perspective of low-SES mothers on contextual factors and information needs and behavior associated with these outcomes.

Methods: A participatory mixed methods research was conducted in partnership with academic researchers and Naître et grandir (N&G) editors. N&G is a magazine, website, and newsletter that offers trustworthy parenting information on child development, education, health, and well-being in a format that is easy to read, listen, or watch. *Quantitative component (QUAN)* included a *3-year longitudinal observational web survey*; participants were mothers of 0- to 8-year-old children. For each N&G newsletter, the participants' perception regarding the outcomes of specific N&G webpages was gathered using a content-validated Information Assessment Method (IAM) questionnaire. Differences between participants of low SES versus others were estimated. *Qualitative component (QUAL) was interpretive*; participants were low-SES mothers. The thematic analysis of interview transcripts identified participants' characteristics and different sources of information depending on information needs. Findings from the two components were integrated (*QUAN+QUAL integration*) through the conceptual framework and assimilated into the description of an ideal-typical mother of low SES (Kate). A narrative describes Kate's perception of the outcomes of web-based parenting information and her perspective on contextual factors, information needs, and behavior associated with these outcomes.

Results: QUAN—a total of 1889 participants completed 2447 IAM responses (50 from mothers of low SES and 2397 from other mothers). N&G information was more likely to help low-SES participants to better understand something, decrease worries, and increase self-confidence in decision making. QUAL—the 40 participants (21 N&G users and 19 nonusers) used 4 information sources in an iterative manner: websites, forums, relatives, and professionals. The integration of QUAN and QUAL findings provides a short narrative, Kate, which summarizes the main findings.

Conclusions: This is the first study comparing perceptions of information outcomes in low-SES mothers with those of other mothers. Findings suggest that equity-oriented, web-based parenting information can offer equal benefits to all, including low-SES

XSL•FO
RenderX

mothers. The short narrative, *Kate*, can be quickly read by decision policy makers, for example, web editors, and might encourage them to reach the underserved and provide and assess trustworthy web-based consumer health information in a format that is easy to read, listen, or watch.

Introduction

Problem and Objectives

This paper explores the outcomes of equity-oriented, web-based, consumer health information from the perspective of young children's mothers who have a low socioeconomic status (SES). Equity-oriented interventions attempt to move toward equity, that is, reduce inequalities (systematic differences among social groups), though full equity may never be achieved [1]. Education and income are among the most important indicators of SES and are strongly associated with individual and population health status [2-4]. With respect to web-based consumer health information, the literature suggests the following stereotypical inequality: although information in general can help people of low SES, it mainly offers benefits to people of higher SES, that is, those who are the most educated and wealthiest. It is unclear whether equity-oriented information, that is, trustworthy information provided in a format that is easy to read, listen, or watch, can offer equal benefits to all, including low-SES mothers.

Regarding general outcomes of web-based information at the population level, the literature suggests mixed positive and negative outcomes. First, health services research suggests that browsing the internet leads to a decrease in unnecessary calls and visits to health professionals and helps to optimize services utilization [5,6]. Specifically, consumers' use of trustworthy sources of web-based health information is associated with improved knowledge, empowerment, self-care, engagement in health care, health outcomes, and quality of life [6-12]. Second, exposure to web-based information can lead to negative outcomes such as worries and anxiety, for example, cyberchondria, deterioration of the patient-provider relationship, and unnecessary visits to the emergency room [13-18].

However, 4 literature reviews have shown that specific outcomes of information are rarely researched [19-22]. Typically, information outcomes are diluted in the outcomes of educational programs and communication with professionals [23-25]. With respect to the health of the mother and child, studies have examined the outcomes of lay pediatric information in the context of the effectiveness of parent-child professional communication and patient education programs. The studies concluded that the parents' level of health literacy is positively associated with the quality of pediatric care, compliance with medical interventions, and child health outcomes but negatively associated with medication errors and visits to emergency rooms [26,27]. This led to the inclusion of the *Low Health Literacy Universal Precaution Principle* in medical education and continuing professional development programs [28-30], which recommends the universal provision of clear lay information in all patient-clinician encounters.

Such principles are essential for web-based consumer health information because the proportion of people with a low level of literacy is substantial; for example, more than 50% of Canadian adults have a low literacy level [31]. People with a low literacy level can read, listen to, watch, and understand one-idea sentences in plain language, but they arguably face difficulties in finding trustworthy web-based information, critically appraising it, and understanding nuanced ideas or specialized language [31]. It follows, therefore, that their web-based health information literacy level is also low, which comprises computer literacy, information literacy, and health literacy. This is particularly important for parents because their low literacy level is detrimental to child health education, healthy behaviors, health, and medication [32-34]. Parents with low education and low income, hereafter referred to as low-SES parents, typically have a low literacy level, yet they have greater information needs compared with their higher-SES counterparts [20].

In the information sciences, we found 5 studies that focused on parenting information needs and information-seeking behavior and included low-SES parents. Two surveys, one in Australia and one in Switzerland, found no statistical difference in internet utilization between low-SES parents and other parents [35,36]. Two qualitative studies on mothers' information needs (mostly middle-class mothers) suggest that practical information on mothering produced by professionals is valuable and helpful [37,38]. One study found that all low-SES mothers search for health information on the internet [39].

In addition, little research has specifically focused on outcomes of web-based parenting information reported by low-SES mothers [40,41]. We found little evidence (and some of it contradictory) on parental perception of outcomes of web-based parenting information, namely information on child education, development, health, and well-being. One study suggested that all parents reported a similar degree of satisfaction with web-based parenting information in general [42]. Five studies suggested that trustworthy web-based information, that is easy to read, listen to, or watch, can improve the quality of life of parents of low SES (including refugee and homeless) and have positive family, economic, and social impacts [43-47]. In contrast, 4 studies showed that barriers persist regarding the acquisition, cognition, and application of web-based parenting information, which are associated with information-related inequalities, perpetuating the digital divide between low-SES and high-SES parents [40,48-50].

However, no study has compared the low-SES parents' perception of outcomes of equity-oriented web-based information with that of other parents. In light of this, this study aims to (1) compare the low-SES mothers' perception of outcomes of web-based parenting information with that of other mothers, and (2) explore the perspective of low-SES mothers on contextual factors and information needs and behavior associated with these outcomes.

Background

This study is based on a partnership between Naître et grandir (N&G) and McGill University (Information Assessment Method, IAM). N&G is funded by the Chagnon Foundation, a philanthropic organization that seeks to prevent poverty. N&G produces a magazine and a website with an email newsletter (in French) for parents of 0- to 8-year-old children.

N&G seeks to raise societal awareness of the importance of early child development for enabling conditions of educational success. It provides free, independent, and trustworthy parenting information to valorize, inform, educate, and equip parents and families of young children, especially among vulnerable populations, for example, tips for parents, dos and don'ts validated by experts. Specifically, N&G provides web-based parenting information content that is easy to read, listen to, and watch, specifically webpages with corresponding videos, podcasts, and computer-audio-assistant highlighting sentences read. The website content is organized by age groups and topics.

N&G is widely read by French-speaking parents across Canada, the United States, and more than 100 other countries. In the 2018 calendar year, 61.6 million N&G webpages were viewed during 35.3 million visits to the N&G website by 15.2 million unique internet protocol (IP) addresses across the world. Among those, 20.2 million webpages were viewed during 11.2 million visits to the N&G website by 3.3 million unique IP addresses in Quebec. More than 213,000 N&G weekly newsletters were emailed to parents in Quebec. *SOM Recherche & Sondages* (personal communication, 2015) conducted a survey of a representative random sample of the population of parents of young children in Quebec, which showed that 82% of respondents knew N&G and that 76% consulted it.

Conceptual Framework

Origin

Our conceptual framework is summarized in Figure 1 and published elsewhere [51]. The framework is based on a mixed studies systematic review with framework synthesis of qualitative and quantitative evidence on the outcomes of web-based consumer health information in community-based primary care. In accordance with Gregor's definition of theories [52], our framework seeks to explain what these outcomes are (five levels of information outcomes), why and how they occur (information needs and seeking behavior), and when, where, and for whom they occur (contextual influencing factors).

Figure 1. Outcomes of web-based parenting information: conceptual framework. SES: socioeconomic status.
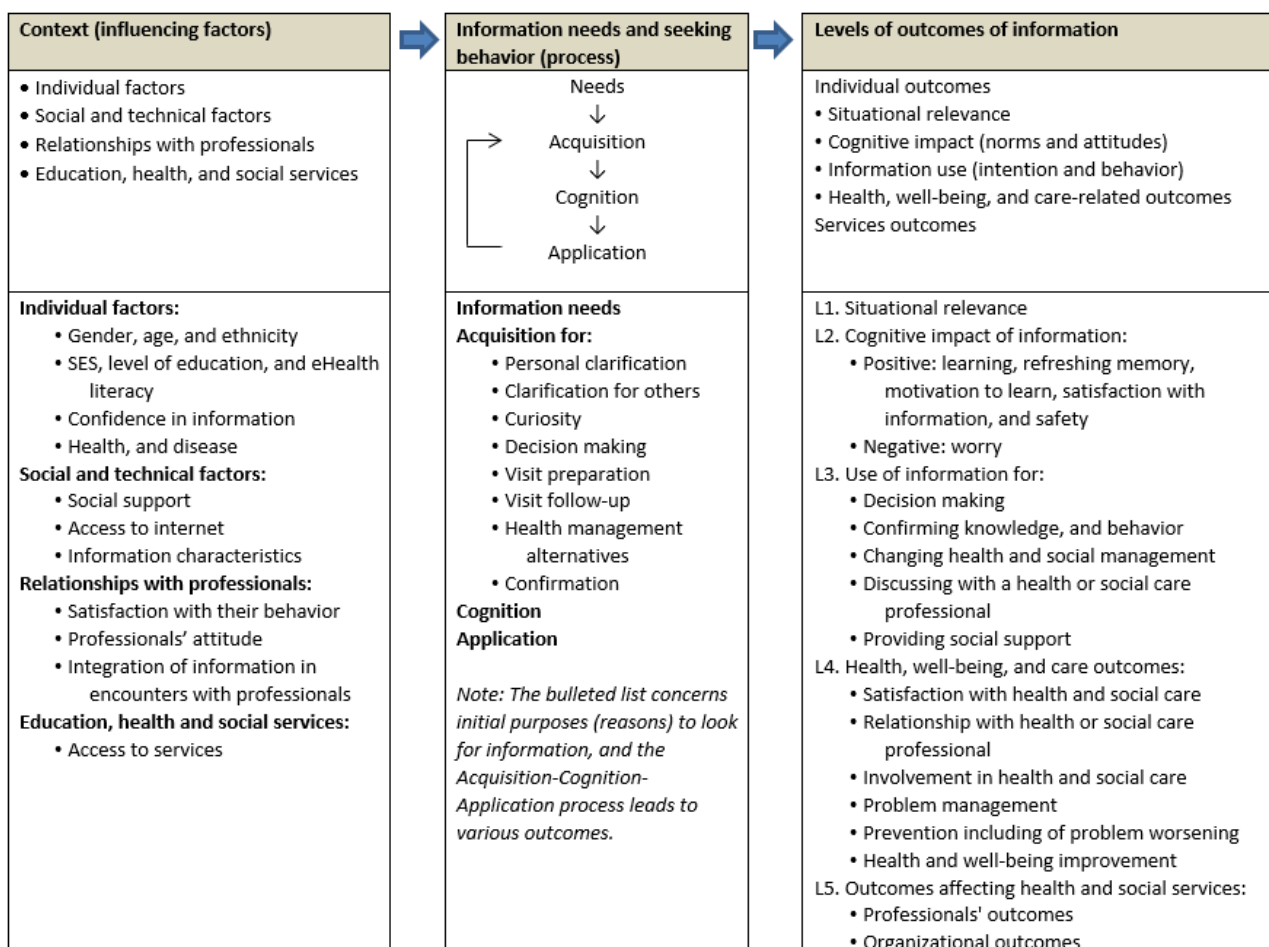
Five Levels of Information Outcomes

The framework includes 4 individual levels of information outcomes (situational relevance, cognitive impact, use, and health and well-being outcomes of information) and one organizational-level outcome (information outcomes affecting educational, health, and social care services).

- Level 1—situational relevance: Relevance of information is the first value of information [53], and situational relevance is the first outcome of information for a person in a specific situation. Parents read or listen to a webpage when information is relevant to their needs, and they skip it when it seems irrelevant.
- Level 2—cognitive impact: Relevant information content has a cognitive impact that parents may perceive as either positive or negative. For example, they can learn either something new from the information they read, hear, or watch, or they may not understand this information. Relevant information with a positive cognitive impact is usually (but not necessarily) used.
- Level 3—information use: Parents can use information content in 4 different ways: conceptual, legitimating, symbolic, and instrumental. For instance, they may use it to decide to consult health professionals (instrumental use) and then share it with a professional (symbolic use). Information use does not necessarily lead to behavior change.
- Level 4—health and well-being outcomes: Information can benefit, or negatively affect, parents' and children's health and well-being, for example, parents might feel reassured (positive outcome) or more anxious (negative outcome) from using information.
- Level 5—organizational outcomes: At this level, parents' information use may influence their utilization of educational, health, and social care services.

Context, Information Needs, and Information-Seeking Behavior

In the framework, information outcomes depend on the context, and parents' information needs and information-seeking behavior. Regarding *context (influencing factors)*, parents' individual and social factors are interrelated because the latter, for example, social support for finding, understanding, and using relevant information, can overcome the former, for example, low level of literacy. These factors are listed in Figure 1. Regarding *information needs and seeking behavior*, outcomes of web-based parenting information and contextual factors are defined in relation to a specific situation: a singular content is sought, acquired, or delivered, for example, a webpage, in a particular situation to address the parents' needs. These conditions are necessary to observe the *imbrication* between information content and information technology and parents—the ultimate decision makers about the value of the information content [54]. In our framework, parents' information-seeking behavior is defined as the iterative imbrication-centered acquisition, cognition, and application of information.

In accordance with the conceptual framework, our specific research questions were as follows:

- Q1: To what extent do mothers of low SES perceive outcomes of N&G web-based parenting information compared with other mothers?
- Q2: What are the experiences of low-SES mothers with web-based parenting information from N&G compared with other sources of information?

Specifically regarding Q2, we wanted to identify the differences between N&G users and nonusers regarding the factors related to information outcomes (contextual factors, and information needs and seeking behavior).

Methods

Approach

We followed an organizational participatory research approach that blends action research and organizational learning to engage organization members to improve practice [55,56]. Researchers and N&G editors partnered throughout the study, developing the research questions together, and making other key research decisions jointly regarding the collection and analysis of data and the interpretation and dissemination of results.

Design

We used a mixed methods convergent design [57,58]. Mixed methods are crucial for *unpacking complexity* in research on poverty and vulnerability [59]. In our study, mixed methods were justified because our quantitative data collected with the McGill IAM questionnaire reflect the importance of the information outcomes perceived by low-SES parents, and qualitative data were required to gain insight into the mechanisms underlying these outcomes. Thus, to respond to our first research question, we conducted a 3-year longitudinal observational quantitative component of the mixed methods study. To answer the second research question, we conducted a qualitative interpretive study on parental characteristics, context, information needs, and information-seeking behavior (the qualitative component of the mixed methods study). Then, we integrated these components via the assimilation of quantitative and qualitative results [58].

Quantitative Longitudinal Observational Component

In line with the literature on patient-reported outcome measures, that is, measurement from patients' perspectives rather than from medical or biological ones [60], we focused on parent-reported information outcomes.

Setting and Participants

People can sign up to receive a weekly newsletter containing links to three N&G webpages tailored to their needs (pregnancy; child 0-1 year, 2-3 years, 4-5 years, and 6-8 years). Participants were recruited from mothers of children 0 to 8 years old from Quebec, identified through the postal code and child's age provided when signing up. They were recruited when they completed at least one IAM questionnaire during the 3-year study period (January 1, 2016, to December 31, 2018, inclusively) and reported an intention to use N&G information for themselves or their child. No incentive was provided to participate.

Hypotheses

According to the literature on the digital divide, we hypothesized that low-SES mothers of 0- to 8-year-old children (ie, mothers with a low level of income and a low level of education) are less likely to perceive positive outcomes from N&G information compared with mothers of higher SES. According to the Quebec poverty line definition, a low level of income is an annual family income of less than Can $40,000 (US $30,070). A low level of education constitutes a high school diploma (grade 12, no university degree) or no diploma (high school not completed).

Instrumentation

The IAM questionnaire was used to assess the information provided on each N&G webpage. The IAM allowed participants to report perceived outcomes associated with information content of a specific webpage in terms of situational relevance, cognitive impact, intention to use, and expected health outcomes [61]. The development of the IAM questionnaire has been based on a theory, literature review, expert panel, and interviews with consumers [62]. N&G editors and McGill researchers have been partnering since 2014 for collecting parents' IAM ratings, and they use IAM-based feedback comments to improve information content (crowdsourcing developmental evaluation) [63].

For each N&G webpage, participants were invited to complete an IAM-parent-v2015 questionnaire by clicking on a lateral tab *survey*. Then, respondents had the opportunity to provide written open-ended feedback about the N&G webpage. To decrease response fatigue, respondents did not receive another invitation during the 30-day period following their completion of the IAM. In preparation for this study, we conducted an ecological content validation of the IAM questionnaire with parents who used the initial version of IAM to rate N&G information [64]. In other words, we questioned N&G IAM respondents regarding the relevance and representativeness of the IAM questions [65,66]. We measured the relevance of IAM questions using the IAM ratings of 22,407 parents, and qualitatively evaluated question representativeness and clarity via interviews with a purposeful sample of 21 parents who used N&G. On the basis of quantitative and qualitative results, McGill researchers and N&G editors revised and clarified the IAM questions. This led to the creation of content-validated IAM-parent-v2015.

Data Source

All participants were asked to complete a demographic questionnaire when they completed their first IAM questionnaire. All IAM and demographic questionnaires were collected by N&G, anonymized (email addresses replaced by an individual identification number), linked, and transferred to a password-protected server for access by the academic members of the research team.

Statistical Analysis

Descriptive statistics were calculated using SAS 9.4 software (SAS Institute). Categorical variables were described as counts and percentages. Ninety-five percent confidence intervals for differences in the proportion of IAM ratings were calculated to estimate outcome differences between the 2 primary study groups: participants combining low level of education and low level of income versus other participants [67,68].

Qualitative Interpretive Component

Reporting

This component is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research [69]. Interviews were conducted with a purposeful sample of 40 low-SES mothers of 0- to 8-year-old children geographically dispersed across Quebec: 21 N&G users and 19 nonusers. Interviews and data analyses were conducted in French. Selected excerpts of transcripts were translated in English for submitting this study to the Journal.

Participants

Women who satisfied the following criteria were purposively recruited through emailed invitations: mother of at least one child aged between 0 and 8 years with a low level of income and a low level of education as defined in the Hypotheses section. We sought to recruit mothers from single- and dual-parent families and living in urban, semiurban, and rural areas. The recruitment strategy involved sending an invitation to mothers who used N&G and had agreed to be contacted for research purposes. Only 3 participants were recruited in this manner. Thus, we pursued recruitment aided by a survey firm. In total, 45 persons were contacted, but 5 were not selected because they did not satisfy the eligibility criteria.

Data Collection

Data collection took place between January and March 2018. Three academic team members (PhD degrees), experts in qualitative research with vulnerable populations (a male anthropologist, a female bioethicist, and a female sociologist with 10-year experience in social work), met participants in either their home or in a research facility room, as per their choice. Individual face-to-face, semistructured interviews of approximately 1-hour duration were conducted. All participants provided formal written consent before the interviews and were compensated for their time following the interview. Participants who were interviewed in their homes received Can $50 (US $38), whereas those who traveled to a research facility received Can $60 (US $45). Interviews were audio-recorded and transcribed verbatim. On the basis of conceptual framework, the research team worked together in an iterative manner to develop the interview guide. Then, the guide was pilot-tested with 2 mothers of young children, a research trainee not working on the project, and a mother of low SES, and revised accordingly. The guide included themes and probes for both N&G users and nonusers. Specifically, all participants were asked about their use of the internet, their information needs, and information-seeking behavior regarding their child's development, education, health, and well-being. The N&G information user participants were asked additional questions about their experience with N&G.

Data Analysis

Data analysis was performed between February 2018 and May 2019. Transcripts were analyzed using NVivo version 11 (NVivo Qualitative Data Analysis Software, QSR International), which helped to maintain a transparent relationship between data (excerpts of transcripts) and themes and subthemes (auditable trail). Two research professionals (interviewers) read transcripts

several times to become fully acquainted with the content before coding the data and used hybrid deductive-inductive thematic analysis [70,71]. An example of the analytical process is presented in Multimedia Appendix 1. They co-constructed a theme tree codebook, using the conceptual framework in a deductive manner, and as new themes and subthemes were found in the data, they were added to this theme tree in an inductive manner. The codebook is presented in Multimedia Appendix 2. All 40 transcripts were analyzed. As shown in Figure 2, saturation of themes and subthemes was reached after 10 coding sessions, that is, once 21 of the 40 interview transcripts had been analyzed. Only 1.5% (5/325) of new themes and subthemes were found in the remaining 19 interviews.

Figure 2. Qualitative data analysis: saturation of themes reached after 10 coding sessions.
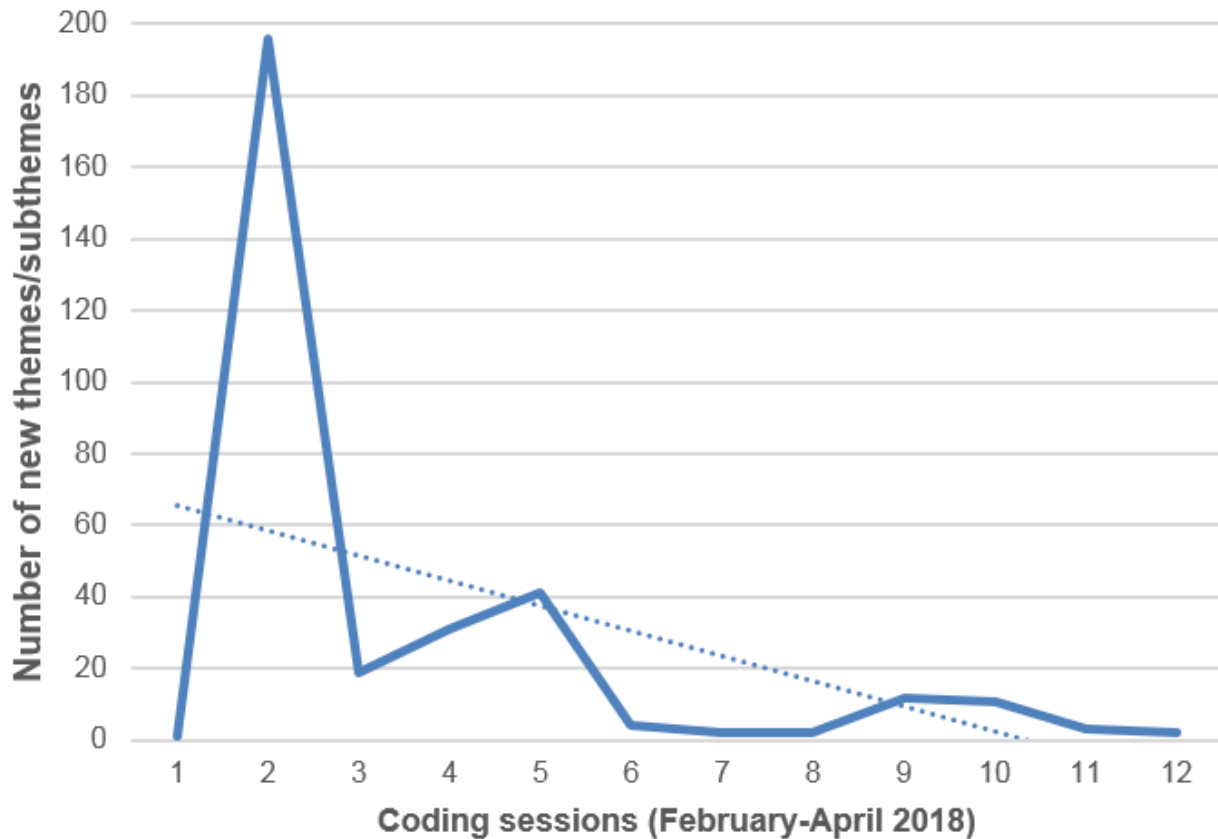
With respect to methodological rigor, data analysis involved multiple sessions where themes, subthemes, and the assignments of transcript excerpts to themes and subthemes were discussed (corroborated or nuanced) among researchers. The reliability of the 2 research professionals conducting the thematic analysis was ensured via the co-construction of the theme tree, initial paired coding sessions on few transcripts, memo writing, and regular meetings with the principal investigators and then with coinvestigators and N&G editors to discuss analysis issues and new themes [70].

Integration of Quantitative and Qualitative Components

The quantitative and qualitative components were integrated in two ways. First, they were integrated using the conceptual framework [57]. Quantitative observations provided results on low-SES mothers' reported outcomes compared with other mothers. Qualitative observations produced complementary results on low-SES mothers' information outcome–related contextual factors, information needs, and information-seeking behavior.

Second, quantitative and qualitative results were assimilated as follows [58]: the principal investigator transformed quantitative results into a narrative, which was combined with an interpretation of qualitative results for building a comprehensive ideal type of low-SES maternal perception of outcomes of web-based parenting information, including outcomes-related factors. All coauthors reviewed the ideal type. According to Weber [72], better understanding can emerge from the scientific construction of an ideal type; here, *ideal* refers to ideas, not to an ideal to achieve. An ideal type does not exist *in the real world under this pure, absolutely ideal form* [72]. It integrates common aspects of a phenomenon in a comprehensive and coherent manner. This integration adds value to qualitative and quantitative results alone.

Results

Quantitative Results

Over the 3-year study period, 1889 participants completed 2447 IAM questionnaires (on average 1.3 questionnaires per participant; range 1-12) pertaining to 683 distinct N&G webpages (on average 3.6 questionnaires per page; range 1-26).

Participants were geographically dispersed across 352 of the 420 Quebec postal codes (352/420, 83.8%). Participants' demographic information, IAM ratings, and comments are detailed below.

Among participants 89.31% (1687/1889) were aged between 25 and 44 years, 94.12% (1778/1889) were living with a partner, 85.23% (1610/1889) were living full time with their child; 11.54% (218/1889) had an annual family income less than Can $40,000 (US $30,070), and 4.71% (89/1889) did not complete high school or had only a high school diploma (grade 12). Compared with the 25- to 64-year-old Quebec population, participants had a higher level of education (Table 1). Compared with couples with children in Quebec, participants had a lower level of income (Table 2). Among participants, 77.77% (1469/1889) provided a valid postal code, and according to the Material and Social Deprivation Index [73], 70.30% (1328/1889) of participants had low levels of education and income and 35.2% (665/1889) of other participants lived in deprived areas.

Table 1. Level of education: participants of the quantitative component versus Quebec 25- to 44-year-old population.

Population	Diploma, n (%)			
	None	Secondary school (grade 11 or less)	College (grade 12 or 13)	University
Participants (n=1889)	21 (1.11)	210 (11.12)	430 (22.76)	1228 (65.01)
Quebec[a] (n=2,056,110)	215,891 (10.50)	709,358 (34.50)	483,186 (23.50)	647,675 (31.50)

[a]Source: Statistics Canada Census of population: Table 37-10-0099-03 distribution of the population of 25 to 64 years old by the highest level of education. Ottawa: Government of Canada; 2016.

Table 2. Level of income: participants of the quantitative component versus couples with children in Quebec.

Population (N; %)	Annual family income, Can $ (US $), n (%)		
	<40,000 (30,070)	40,000-80,000 (30,070-60,140)	80,000 (60,140)
Participants (n=1889)	217 (11.49)	586 (31.02)	1086 (57.49)
Quebec[a] (n=911,975)	49,835 (5.47)	319,170 (35.00)	542,970 (59.54)

[a]Source: Statistics Canada Census of population: Catalog no. 98-400-x2016129. Ottawa: Government of Canada; 2016.

The types of information outcomes are detailed in Table 3. Specifically, all ratings (2447/2447, 100.00%) reported the participants' intention to use N&G information for themselves or for their child, at least in a conceptual manner. In 97.79% (2393/2447) of ratings, participants expected that information use would lead to health and well-being benefits for themselves or their child.

Figure 3 displays the estimated group differences in proportions for these outcomes, along with 95% CI. The comparison of the different types of intention to use and expected health and well-being benefits of N&G information between the IAM ratings of the participants with low levels of education and income (n=50) and those of the other participants (n=2397), suggested that 5 of 11 outcomes were not inferior for the former group (items a, d, e, g, and k in Figure 3), including 3 being superior (items a, g, and k in Figure 3). The results do not provide conclusive evidence for differences (or the absence of such differences) regarding the 6 other types of outcomes (items b, c, f, h, i, and j in Figure 3). In other words, N&G information was more likely to help participants with low levels of education and income to better understand something (mean difference 0.144; 95% CI 0.007-0.253), decrease worries (mean difference 0.171; 95% CI 0.036-0.302), and increase self-confidence in making a decision with someone else (mean difference 0.115; 95% CI 0.015-0.249).

Table 3. Quantitative component—perceived information outcomes: Information Assessment Method ratings of participants with a low level of education and income versus other participants.

IAM[a] questions and response options	Low-education and low-income mothers (50 IAM ratings), n (%)	Other mothers (2397 IAM ratings), n (%)	All participants (2447 IAM ratings), n (%)
Q1. Is this information relevant?			
Very relevant (this is the information I expected)	32 (64)	1629 (67.96)	1661 (67.88)
Relevant	17 (34)	748 (31.21)	765 (31.26)
Somewhat relevant	0 (0)	17 (0.71)	17 (0.69)
Very little relevant (this is not the information I expected)	1 (2)	3 (0.13)	4 (0.16)
Q2. Did you understand this information?			
Very well (I understood everything)	45(90)	2243 (93.58)	2288 (93.50)
Well	5 (10)	152 (6.34)	157 (6.42)
Poorly	0 (0)	1 (0.04)	1 (0.04)
Very poorly (I did not understand much)	0 (0)	1 (0.04)	1 (0.04)
Q3. What do you think about this information? Check all that apply.			
This information allowed me to validate what I do or did	30 (60)	1655 (69.04)	1685 (68.86)
This information taught me something new	22 (44)	1207 (50.35)	1229 (50.22)
This information reassured me	23 (46)	995 (41.51)	1018 (41.60)
This information refreshed my memory	12 (24)	787 (32.83)	799 (32.65)
This information motivated me to learn more	17 (34)	472 (19.69)	489 (19.98)
There is a problem with this information	1 (2)	28 (1.17)	29 (1.19)
I disagree with this information	0 (0)	13 (0.54)	13 (0.53)
This information can have negative consequences	0 (0)	5 (0.21)	5 (0.20)
Q4. Will you use this information?			
Yes	50 (100)	2397 (100.00)	2447 (100.00)
No	0 (0)	0 (0)	0 (0)
Q4a. How you will use this information for you or your child? Check all that apply.			
This information will help me to better understand.	36 (72)	1345 (56.11)	1381 (56.44)
I will use this information to do something in a different manner.	9 (18)	584 (24.36)	593 (24.23)
I will use this information to discuss with someone else.	9 (18)	543 (22.65)	552 (22.56)
I did not know what to do, and this information will help me to do something.	11 (22)	356 (14.85)	367 (15.00)
I knew what to do, and this information convinced me to do it.	5 (10)	123 (5.13)	128 (5.23)
Q5. Do you expect any benefit for you and your child from using this information? Check all that apply.			
This information will help me to improve the health or well-being of my child.	31 (62)	1603 (66.88)	1634 (66.78)
This information will help me to be less worried.	27 (54)	858 (35.79)	885 (36.17)
This information will help me to prevent a problem or the worsening of a problem.	12 (24)	676 (28.20)	688 (28.12)
This information will help me to handle a problem.	14 (28)	615 (25.66)	629 (25.70)
I will be better prepared to discuss with someone else.	11 (22)	547 (22.82)	558 (22.80)
I will be more confident to decide something with someone else.	13 (26)	335 (13.98)	348 (14.22)
I expect no benefits.	1 (2)	53 (2.21)	54 (2.21)

[a]IAM: Information Assessment Method.

Figure 3. Types of information outcome: differences in proportions comparing the Information Assessment Method ratings of participants combining low education and low income versus other participants.
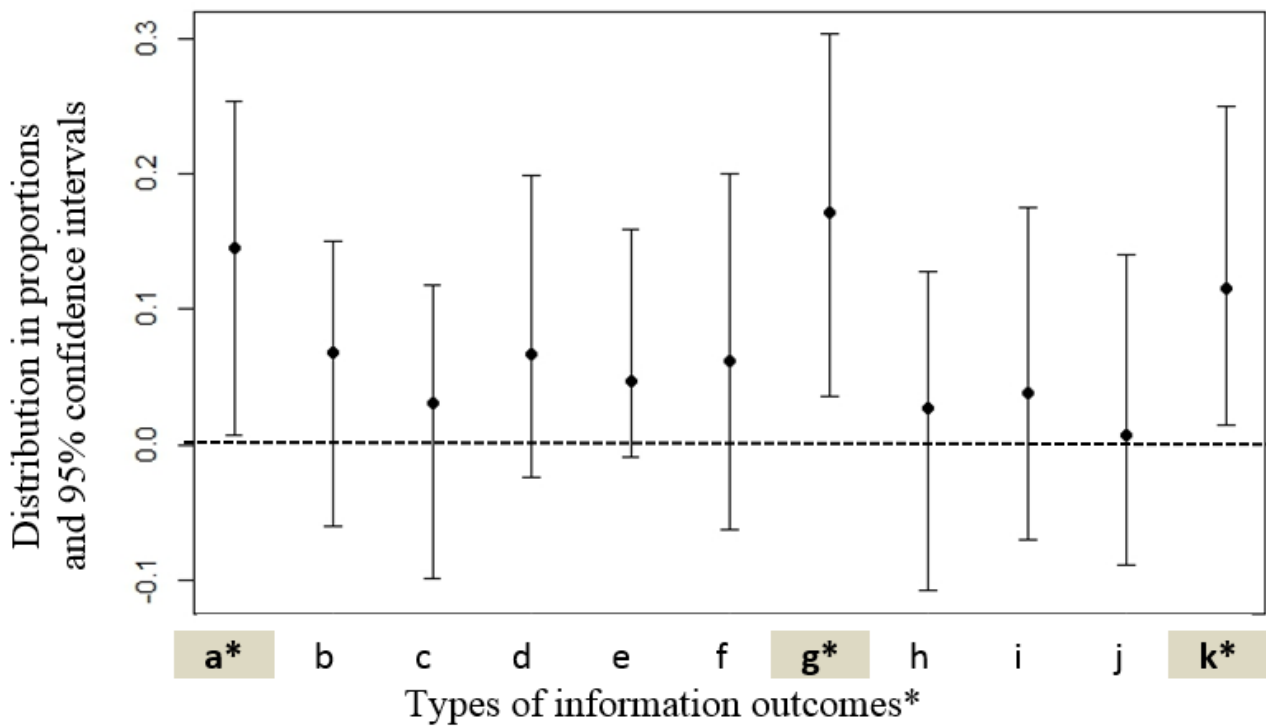
Qualitative Results

Participants' Characteristics

The participant characteristics are described in Table 4. Forty low-SES mothers with at least one child aged between 0 and 8 years were interviewed: 21 N&G information users and 19 nonusers. On average, they had 2 children (range, 1–5 years; median, 2), including at least one child aged 0 to 8 years (mean 5.9 years, range pregnancy to 8 years, and median 4 years). Interviews revealed that about half were just above the poverty line (eg, Can $40,000-50,000 [US $30,070-38,085] per year) or the low level of education (eg, professional certificate).

The results reveal no major differences between N&G information users and nonusers with respect to the frequency of internet use and the type of information retrieved. Participants mentioned child education, entertainment, goods, health, news, recipes, social media, and transportation regardless of whether (1) they used N&G information or not; (2) where and when the internet is used at home, public, or work setting, anytime but mostly in the evening; and (3) the languages in which the internet is used, that is, mostly French, English, and Spanish.

Table 4. Qualitative component: sociodemographic characteristics of the participants (interviewees).

Characteristics	Participants, n (%)
Utilization of N&G[a]	
Users	21 (53)
Nonusers	19 (48)
Language spoken at home	
French	27 (68)
French and English	3 (8)
French and Spanish	2 (5)
Spanish	3 (8)
French and Romanian	1 (3)
Unknown	4 (10)
Education	
General education (grade 11 or less)	16 (40)
Professional education (grade 9, 10 or 11)	8 (20)
College (grade 12 and 13 in-progress or completed)	16 (40)
University	0 (0)
Annual family income, Can $ (US $)	
Less than 10,000 (7917)	0 (0)
10,000-20,000 (7917-15,234)	5 (13)
20,000-30,000 (15,234-22,851)	14 (35)
30,000-40,000 (22,851-30,070)	19 (48)
40,000-50,000 (30,070-38,085)	2 (5)
Over 50,000 (38,085)	0 (0)
Parental status	
Single mother	16 (40)
Couple	23 (58)
Not specified	1 (3)
Housing	
Rent	29 (73)
Own	7 (18)
Not specified	4 (10)

[a]N&G: Naître et grandir.

Participants' Informational Context

The contextual factors were as follows. Regarding individual characteristics, on average, participants were 32 years old; 24 were seeking a preuniversity degree equivalent to 2 years of general or professional education after grade 11, and the other 16 had stopped their education after grade 11. Interviews analysis revealed that 16 participants were single mothers who earned less than Can $25,000 (US $19,043) per year, 22 other participants had an annual family income of less than Can $40,000 (US $30,070), and 2 participants earned between Can $40,000 and Can $50,000 (between US $30,070 and US $38,085). Furthermore, 18 participants mentioned that one of their children had a health or developmental problem (12/21

N&G information users vs 6/19 nonusers): attention deficit hyperactivity disorder (n=6), language delay (n=6), cardiac problem (n=2), autism spectrum disorder (n=2), and growth delay (n=2).

As illustrated below, all participants stated that they used the internet several times per day with their cellular phone. Almost all participants considered it easy to access the Web (17/21 N&G information users vs 19/19 nonusers), given, for example, the presence of free Wi-Fi at work, and in numerous public places (shops, schools, etc). All participants reported using their cellular phone to access the internet, whereas only 5 participants used computers or tablets. Participants said they used their cellular phone more frequently because it is more portable and

personal, and because it is accessible at all times, unlike a tablet or computer or that can be shared with other family members and, in the case of computers, needs to be booted up. Only 5 participants mentioned barriers to internet access, namely the cost of cell phone data. Two mentioned they had to reduce their data consumption as it was too expensive, and thus accessed the internet using free Wi-Fi when possible:

> *As soon as you ask yourself a question, you just have to go on the Internet, and you'll get your answer. Now, it seems like we're not used to not knowing something. [...] Right now, you take out your phone, you know it right away [laughs].* [D01]

> *Even before, I didn't pay to have it [Internet] on my phone, I just had it at home or in places with Wi Fi; it was pretty much just at home that I used it, but now*

> *I have it everywhere I want, well... We can share our connection.* [D12]

Participants' Information Needs and Seeking Behavior

Information Sources

We found few differences between N&G information users and nonusers regarding information needs specific to child development (19/21 N&G users vs 9/19 nonusers) and education (7/21 N&G users vs 1/19 nonusers). Participants identified 4 sources accessed for information seeking: websites, web-based forums, relatives, and care services (no particular difference between users and nonusers). These sources are summarized in Table 5. Depending on information needs, participants described searching multiple sources of information in an iterative manner (including magazines and books from the library, radio, and television).

Table 5. Qualitative component: participants' information-seeking behavior.

Descriptor	Parenting information-seeking behavior			
Sources[a]	Websites: experts providing trustworthy information on websites (eg, N&G[b])	Web-based forums: credible people with life experience similar to the participant's experience	Relatives: trusted relatives with children (eg, mothers, sisters, and friends)	Care services: trustworthy professionals from health and social care services[c]
Why	• Nonurgent problem • Easy (24/7) • Autonomy	• Nonurgent problem • Similar values • Breaking isolation	• Nonurgent problem • Affective support • Mutual understanding	• Urgent or severe issue • High risk • Uncertainty
Facilitators	• Easy-to-read, listen to, and watch information • Corroboration by different websites	• Comparison of information content with other sources	• Routine information exchanges • Knowledge of children of participants	• Free easy access to health and social care services • Reassurance
Barriers	• No lay information • Lack of content on child development and education (compared with health)	• Lack of quality control • Incivilities • Lack of content on child development and education	• Different child-related values and preferences (eg, generation gap)	• Difficulties to navigate services, specifically services addressing behavioral and social issues

[a]The 40 participants address their information needs via one or more of the following sources.

[b]N&G: Naître et grandir.

[c]Including professionals and staff from community organizations.

Source 1–Websites

Typically, participants referred to this as *Googling*. Websites without social interaction, such as N&G, were the most common type of first-contact source, especially when seeking information in a nonemergency situation. Governmental and institutional websites were mentioned as sources of trustworthy information. Websites were considered easy to navigate and convenient (available 24/7). Participants reported that they provide a variety of information content on numerous topics and languages, which allows them a degree of autonomy and independence. However, they mentioned that information on child development and education is more difficult to retrieve than general health information, whereas this is not the case in N&G:

> *What I often do, since I speak all three languages, Spanish, French, English, I go and search [online] in all three languages. So, we often go and see different perspectives of the subject, and then I make*

> *my mind up about it. [...] That way I have a more global idea.* [D10]

> *For me it's easy because I know how to navigate, I know how to search for information [online]; [...] For example, N&G tab "Age 5-6": I go there. It's all there. It's fast. When you read, for example "at this age, they do role-playing games," you don't know what role-playing games are, [...] you click on it and it takes you to another page that explains what role-playing games are. [...] I find it very practical.* [E91]

> *Sure, if it's anything medical, I know where to go. If it's dental, I know where to go. But like I think, maybe on a more behavioral level, then I'm like [not knowing where to go].* [E01]

Source 2–Web-Based Forums

Participants considered web-based forums an interesting source for information on general issues, as they provide information from people who have experienced a similar issue, and thus considered credible. Participants sought out these forums for isolation-breaking interactions, specifically looking for people with similar values with whom to interact. They compared the informational content from forums with other sources of information when questioning information quality, for example, the presence of offensive language and inappropriate comments in a forum. As with websites, participants reported that web-based forums lack content on child development and education:

> I'm pretty isolated here. My husband, his family is here, so that's why we came here, but otherwise, […] I have no social circle and no family here. So [online forums] make me socialize a little bit with women that I might have some affinity with. [E09]

> There are 800 of us in that group [online], so there are 800 people who can give me an answer. What's also good is that there are all kinds of specialties, pharmacists, nurses, physiotherapists, chiropractors... Depending on the question, there's definitely someone specialized who can answer me in addition to sharing their experience as a mom. [D12]

> I'm on a mom forum on Facebook, it's like a private group. So the first thing I usually do is I ask for advice, and that's for those who have children of the same age. Then I do a little research [online] to compare. [E12]

Source 3–Relatives

Participants reported that parents and relatives with children constituted a trusted source of information on child development, education, and nonemergency situations. These individuals provided effective support and shared familiar experiences. Mutual understanding between relatives and participants appears to be enabled by a history of information exchanges and relatives' knowledge of the participants' children. Not all participants had relatives with children, and some reported different values and preferences compared with their family members, which limited this source:

> My first instinct is always to consult someone close to me. It really depends on the moment. If there is no rush and if the person is available, I call on my relatives. [D18]

> I consult friends who have children, and lastly my mother [laughs]. [D19]

> I often call my mother to find out. But it's not always good what parents tell you. They often share old wives' tales, and in their day it worked like that. But now the products have changed and often what my mother tells me is not good. [T05]

Source 4–Care Services

All participants trusted information provided by professionals working in health and social care services. For example, they never hesitated to call the Quebec HealthInfo 24/7 free phone line or to visit a clinical setting. They used care services when faced with an emergency, such as an accident, perceived potential risks associated with self-care, and uncertainty, for example, hesitant to make a self-diagnosis or to use a homemade remedy. Participants praised easy access to public services and community organizations, and reassurance (decrease worries) obtained from professionals. In turn, this appears to enable and reinforce the use of this source. However, almost all participants reported difficulties in finding information about, and navigating, social care services:

> You can ask questions [to pharmacists], and you won't have to wait seven or eight hours in the emergency room. [D17]

> My young child has a stye right now. I know for a fact that it is a stye, but I didn't take a chance and I went to the doctor for the doctor to tell me "yes, that's what he has, then yes, you're doing the right thing, keep doing the compresses and that's it." I knew what I had to do. I've been through this before because my [other] child is 17 and had a stye, so I know what it is, but I need to be reassured. [E03]

> If [my question] is a medical question, I know where to go. […] At the behavioral level, let's say I always tend to think: Okay, but who am I going to see? [E01]

Searching for Multiple Sources of Information in an Iterative Manner

Typically, the participants' information-seeking behavior consists of an iterative process that includes multiple sources of information. For instance, some participants reported beginning by looking for information on websites because it is convenient. They then consult web-based forums and relatives to search for complementary and corroborating information. Finally, they consult with health and social care services to confirm what they have found or to reduce uncertainty when they find contradictory content from websites, forums, and relatives. Some participants reported that they started their information search by consulting relatives and then sought to corroborate and confirm information obtained using websites, web-based forums, and care services. Participants mentioned that they consult care services eg, HealthInfo 24/7 phone line, directly when their information needs pertain to a perceived emergency. They may then seek corroboration and additional information through websites, forums, and relatives:

> I find it easier today with the internet. You have a lot of accessibility. When I have a question, I'll turn to my sisters, my mom, but if they're not available, sometimes I'll just Google my question, and I'll find forums. [T07]

> The first reflex is always to consult someone close to you. That depends on the moment. If the question is urgent and we don't have [access] to the person, we take what we have at hand, the Internet. [D18]

> When they have a fever, or are sick, or are in pain somewhere, then I'll go and look at all kinds of places.

[...] Calling at places sometimes you can't get through right away. [D15]

Integration of Quantitative and Qualitative Results: The "Kate" Ideal Type

The ideal type, Kate, is a 30-year-old single mother with 2 children who are 2 and 6 years old. She has completed secondary school and is interested in acquiring professional education. She accesses the internet multiple times daily using her cellular phone. Her annual income is Can $20,000 (US $15,234), and thus she cannot afford additional internet fees to use on a tablet or computer. However, she has easy access to the internet when needed by using her neighbors' and relatives' Wi-Fi connection, with their permission, and by using free Wi-Fi services in the mall, a local community organization, and the public library. She says, "I take my phone everywhere and can be on the internet almost any time. It's mine and is my link to friends, parents, and the world." When she wants to know something about anything, her first reflex is to look at her phone.

When she needs information regarding the education, health, and well-being of her children, she starts by browsing the top-listed websites on an internet search engine. She compares webpages on a topic from multiple websites, seeking coherence and corroborating information, and confirms her findings using trustworthy governmental and noncommercial websites such as N&G. Usually, she finds relevant and understandable information on N&G that supports what she is planning to do or teaches her something new. Using N&G informational content for her children often decreases her worries and leads her to expect improvement regarding her children's health or well-being and in her interaction with health or social care professionals. Specifically, she finds that N&G provides valuable information about child behavior, development, and education.

In addition, she seeks complementary information to her *Mom-like-me* social media group, specifically by a group member tagged as a mother and a nurse. Concomitantly, she calls her mother and her best friend, who has children of the same age as hers. In this way, she receives valuable reassurance and emotional support from relatives who know her and her children well. Finally, she considers all of the information she has obtained and filters it through her values, preferences, and financial capabilities to make child-related decisions. She calls the HealthInfo 24/7 phone line or consults with health professionals in the case of a perceived emergency, when she has doubts regarding the information she has obtained, or when she faces contradictory information and uncertainty.

Discussion

Principal Findings

Results suggest that in families of young children, there are few differences between low-SES mothers compared with higher-SES mothers with respect to the perceived benefits of equity-oriented, that is, trustworthy and easy-to-read, listen to, and watch N&G web-based parenting information (situational relevance, cognitive impact, use, and subsequent health and well-being outcomes of information). According to our

qualitative results, this hypothesis is limited to nonurgent problems because participants primarily turn to professionals for emergency and risk-related information needs. This hypothesis can be tested in future research with other websites using the web version of IAM, or the mobile version for smartphones released in January 2019.

Specifically, these results do not support our literature-based initial hypothesis that participants' level of education and income would be associated with information benefits. Our results suggest the following hypotheses for future research. Compared with other mothers, low-SES mothers might be more likely to report that web-based parenting information that is trustworthy and easy to read, listen to, and watch helps them to better understand child-related issues and to be less worried and more confident in making decisions with someone else, for example, a professional. They may also report that web-based information helps them to discuss issues with someone else, for example, a relative, and do things differently, better manage or prevent a child behavior problem, improve their health or the health of their child, and be satisfied with professional services.

The results from this evaluation of specific N&G information webpages advance the knowledge on information outcomes. Although mixed methods in research on poverty are expanding rapidly [59], our mixed methods study is the first to compare the outcomes of web-based parenting information from the perspective of low-SES mothers seeking specific child-related information versus other mothers. In this study, the advantages of mixed methods are illustrated by the synergy between our quantitative results on how often information outcomes are reported by low-SES mothers compared with mothers of higher SES and our qualitative results on why and how low SES mothers retrieve and corroborate information.

The low-SES mothers we interviewed frequently access the internet. Moreover, the results of a recent survey of a representative sample of the New York State population (n=1350) suggested that the level of health literacy is not associated with the volume of internet utilization [74]. The low-SES mothers we interviewed considered websites and web-based forums as information sources that are helpful in everyday life. Some relied on their social network, for example, relatives, to obtain information and emotional support before seeking information on websites and web-based forums and from care services. This is congruent with the literature [75-79].

Furthermore, our participants might have been atypical for two reasons. First, they volunteered and, thus, may be biased in support of N&G. Second, although McCloud et al [48] suggested that people of low SES in the United States may experience difficulties accessing the internet despite technical support (provision of computers, home internet connection, and technician help), all of our participants reported that they were able to access the internet easily and none reported technical barriers.

Our study has three main limitations. First, for the quantitative component, we used a convenience sampling strategy (self-selected volunteer participants), and participants did not complete an IAM questionnaire each time they visited the N&G website. This limitation probably led to an overestimation of

positive outcomes (social desirability bias). However, assuming this bias influenced all participants in the same manner, regardless of their levels of education and income, this limitation did not affect the statistical analysis. Second, for the qualitative component, about half of the participants were just above the poverty line or the low level of education threshold, which certainly influenced the results. Third, parental efficacy was not examined in our interviews, but this might have affected our data, for example, information-seeking behavior. Numerous studies have shown that poverty has a detrimental effect on parenting, including perceived parental efficacy [80-82]. The concept of perceived parental efficacy is defined as "beliefs or judgments a parent holds on their capabilities to organize and execute a set of tasks related to parenting a child" [83].

In contrast to general health information, our results show that participants face difficulties in finding information on child development and education on websites (other than N&G) and web-based forums and navigating health and social care services when they have nonmedical care needs. In line with patient-centered care [84], future research may address several questions regarding child behavioral, developmental, and educational issues. To what extent can parents retrieve easy-to-read, listen to, and watch information in this field when needed? In what ways do websites and forums provide sufficient information to guide parents in managing these issues on their own when they desire and are able to do so? How do parents know who to ask for help when this information is insufficient? How can better understand what social care services are available and how they can be accessed?

Our results support approaches for improving information exchange between low-SES mothers and health or social care professionals, including (1) prescriptions of information that is considered trustworthy, easy to read, listen to, and watch; (2) information seeking by trusted relatives who have experience and higher level of eHealth literacy; and (3) referral to another professional, third party, when needed [84,85]. In other words, professionals can help low-SES mothers find, understand, evaluate, and use information because the provision of informational content alone does not usually satisfy the constraints of low education and empowerment among people of low SES [86].

Finally, the "Kate" narrative can be quickly read by decision policy makers, for example, web editors, and might encourage them to reach the underserved. In line with Reichwein et al [87], an ideal type can raise awareness and allow information providers, for example, web editors, to (1) tailor information content for the information needs and seeking behavior of specific target audience, such as low-SES mothers; (2) promote facilitators and overcome barriers to optimize information outcomes in the targeted low-SES audience; and (3) avoid stigmatization of people of low SES by openly reaching everybody (universalism).

Considering that narrative results of evaluations of interventions that show an impact on inequalities are especially useful for policy making [1], the "Kate" ideal type suggests the following main messages and policy recommendations. Web editors and experts can be encouraged to assess how valuable the informational content they share is from the perspective of low-SES people, providing thereby more equity-oriented content (easier-to-read, listen to, and watch information). In line with Luhmann theory of communication-based autopoietic self-referential interrelated social systems [88-91], the *Low Literacy Universal Precaution Principle* must be applied to enable trust and satisfactory communication (information exchanges) between Kate and a variety of social systems, such as education, health, and social services. People rely on website information content when they understand and trust it [92,93], whereas they do not use the content when they do not trust it, and often stop accessing it altogether [94]. In addition, O'Neill [95] specified that misplaced mistrust can be harmful. This applies to our work as follows: information providers and decision policy makers should promote trust in trustworthy information and should furnish resistance to mistrust in trustworthy information and trust in untrustworthy information.

In other words, information providers, such as web editors, must be encouraged, trained, and supported to provide trustworthy information using a plain language standard (readability grade 3-5), audio-assistance, and visuals [96]. This is important as a systematic review of 157 cross-sectional studies that assessed 7891 websites with consumer health information showed that the mean readability grade level ranged from grades 10 to 15 [96]. Thus, people of low SES and a low level of literacy are disproportionately disadvantaged regarding web-based information sources (readability grade superior to 5). Specifically, such inequalities remain with most governmental websites [97].

Conclusions

This study suggests a main message, a feasible intervention, and policy recommendations that can be implemented across health care systems.

Message

Mixing quantitative and qualitative methods and results, this study shows that the perceived benefits from trustworthy easy-to-read, listen to, and watch web-based parenting information are higher for low SES mothers of young children compared with other mothers.

Intervention

Quantitative results are based on the IAM, a method that may be of interest to information providers. Specifically, assessing perceived outcomes of equity-oriented web-based information can help to verify that all people can obtain information they need, thus satisfying the democratic right to information [98-100].

Policies

This concurs to state that literacy-related education courses and continuing professional development activities are necessary across all health and social care disciplines. Our paper concludes with a call for applying universal plain language standards to all governmental and public websites, in an effort to reduce information-related inequalities and to humanize the development of communication technologies and the virtual world.

Multimedia Appendix 1
Example of the analysis process: from excerpts of interviews to themes.
[PDF File (Adobe PDF File), 102 KB - jmir_v22i11e22440_app1.pdf]

Multimedia Appendix 2
Codebook: list of themes and subthemes (NVivo report).
[PDF File (Adobe PDF File), 154 KB - jmir_v22i11e22440_app2.pdf]

References

1. Farrer L, Marinetti C, Cavaco YK, Costongs C. Advocacy for health equity: a synthesis review. Milbank Q 2015 Jun;93(2):392-437 [FREE Full text] [doi: 10.1111/1468-0009.12112] [Medline: 26044634]
2. Baker DP, Leon J, Smith Greenaway EG, Collins J, Movit M. The education effect on population health: a reassessment. Popul Dev Rev 2011;37(2):307-332 [FREE Full text] [doi: 10.1111/j.1728-4457.2011.00412.x] [Medline: 21984851]
3. Mikkonen J, Raphael D. Social Determinants of Health: Canadian Perspectives. Toronto, CA: York University School of Health Policy and Management; 2010.
4. Mechanic D. Population health: challenges for science and society. Milbank Q 2007 Sep;85(3):533-559 [FREE Full text] [doi: 10.1111/j.1468-0009.2007.00498.x] [Medline: 17718667]
5. Korenstein D, Falk R, Howell EA, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. Arch Intern Med 2012 Jan 23;172(2):171-178. [doi: 10.1001/archinternmed.2011.772] [Medline: 22271125]
6. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
7. Smith S, Duman M. The state of consumer health information: an overview. Health Info Libr J 2009 Dec;26(4):260-278 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00870.x] [Medline: 19930474]
8. Prescott J, Mackie L. 'You sort of go down a rabbit hole...You're just going to keep on searching': a qualitative study of searching online for pregnancy-related information during pregnancy. J Med Internet Res 2017 Jun 5;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302] [Medline: 28583906]
9. Erdem SA, Harrison-Walker LJ. The role of the internet in physician–patient relationships: the issue of trust. Bus Horiz 2006 Sep;49(5):387-393. [doi: 10.1016/j.bushor.2006.01.003]
10. Edwards M, Davies M, Edwards A. What are the external influences on information exchange and shared decision-making in healthcare consultations: a meta-synthesis of the literature. Patient Educ Couns 2009 Apr;75(1):37-52. [doi: 10.1016/j.pec.2008.09.025] [Medline: 19036550]
11. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007 Jul 23;167(14):1503-1509. [doi: 10.1001/archinte.167.14.1503] [Medline: 17646604]
12. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US national health interview survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
13. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010 Feb 28;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
14. Lauckner C, Hsieh G. The Presentation of Health-related Search Results and Its Impact on Negative Emotional Outcomes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; May 2, 2013; Paris, France. [doi: 10.1145/2470654.2470702]
15. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009 Nov 1;27(4):1-37. [doi: 10.1145/1629096.1629101]
16. Markoff J. Microsoft Examines Causes of 'Cyberchondria'. The New York Times. URL: http://www.nytimes.com/2008/11/25/technology/internet/25symptoms.html?_r=0 [accessed 2020-02-15]
17. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord 2014 Mar;28(2):259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]

18.   El Sherif R, Pluye P, Thoër C, Rodriguez C. Reducing negative outcomes of online consumer health information: qualitative interpretive study with clinicians, librarians, and consumers. J Med Internet Res 2018 May 4;20(5):e169 [FREE Full text] [doi: 10.2196/jmir.9326] [Medline: 29728350]

19.   Case DO, O'Connor LG. What's the use? Measuring the frequency of studies of information outcomes. J Assn Inf Sci Tec 2015 Jan 30;67(3):649-661. [doi: 10.1002/asi.23411]

20.   Case D, Given L. Looking for Information: a Survey of Research on Information Seeking, Needs, and Behavior. Fourth Edition. London, UK: Academic Press; 2016.

21.   Robson A, Robinson L. The information seeking and communication model. J Doc 2015 Sep 14;71(5):1043-1069. [doi: 10.1108/jd-01-2015-0023]

22.   Urquhart C, Turner J. Reflections on the value and impact of library and information services. Perform Measure Metric 2016 Apr 11;17(1):5-28. [doi: 10.1108/pmm-01-2016-0001]

23.   Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education. Fifth Edition. San Francisco, CA: Jossey-Bass; 2015.

24.   Richard C, Lussier MT. La communication professionnelle en santé. Montréal: Pearson ERPI; 2016.

25.   Thoër C, Lévy J. Internet et santé. Montréal: Presses de l'Université du Québec; 2012:2760535223.

26.   Connelly R, Speer M. Health literacy and health communication. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes. Cham: Springer; 2017:1-13.

27.   Speer M. Health literacy and child health outcomes: from prenatal to birth and infant stages. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes. Cham: Springer; 2017:15-18.

28.   Connelly R, Gupta A. Health literacy universal precaution strategies for communication with all patients. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes. Cham: Springer; 2017:39-50.

29.   Gupta A, Speer M. Health literacy effective health communication in pediatric practices health systems: creating shame-free environments patient-friendly institutions. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes. Cham: Springer; 2017:51-56.

30.   Turner T. Health literacy medical education. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes. Cham: Springer; 2017:57-72.

31.   Ronson McNichol B, Rootman I. Literacy and health literacy: new understandings about their impact on health. In: Raphael D, editor. Social Determinants of Health: Canadian Perspectives. Toronto: Canadian Scholars Press; 2016:261-290.

32.   Weightman AL, Morgan HE, Shepherd MA, Kitcher H, Roberts C, Dunstan FD. Social inequality and infant health in the UK: systematic review and meta-analyses. BMJ Open 2012;2(3):-. [doi: 10.1136/bmjopen-2012-000964] [Medline: 22700833]

33.   Easton P, Entwistle VA, Williams B. Health in the 'hidden population' of people with low literacy. A systematic review of the literature. BMC Public Health 2010 Aug 5;10:459 [FREE Full text] [doi: 10.1186/1471-2458-10-459] [Medline: 20687946]

34.   Connelly R, Turner T. Health literacy and child health outcomes. Champringer. ISBN 2017:3319507990. [doi: 10.1007/978-3-319-50799-6]

35.   Baker S, Sanders MR, Morawska A. Who uses online parenting support? A cross-sectional survey exploring Australian parents' internet use for parenting. J Child Fam Stud 2016 Nov 14;26(3):916-927. [doi: 10.1007/s10826-016-0608-1]

36.   Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]

37.   Carolan M. Health literacy and the information needs and dilemmas of first-time mothers over 35 years. J Clin Nurs 2007 Jun;16(6):1162-1172. [doi: 10.1111/j.1365-2702.2007.01600.x] [Medline: 17518891]

38.   Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]

39.   Guendelman S, Broderick A, Mlo H, Gemmill A, Lindeman D. Listening to communities: mixed-method study of the engagement of disadvantaged mothers and pregnant women with digital health technologies. J Med Internet Res 2017 Jul 5;19(7):e240 [FREE Full text] [doi: 10.2196/jmir.7736] [Medline: 28679489]

40.   Dworkin J, Connell J, Doty J. A literature review of parents' online behavior. Cyberpsychology 2013 Jul 1;7(2):-. [doi: 10.5817/cp2013-2-2]

41.   Moon RY, Mathews A, Oden R, Carlin R. Mothers' perceptions of the internet and social media as sources of parenting and health information: qualitative study. J Med Internet Res 2019 Jul 9;21(7):e14289 [FREE Full text] [doi: 10.2196/14289] [Medline: 31290403]

42.   Rothbaum F, Martland N, Jannsen JB. Parents' reliance on the Web to find information about children and families: socio-economic differences in use, skills and satisfaction. J Appl Dev Psychol 2008 Mar;29(2):118-128 [FREE Full text] [doi: 10.1016/j.appdev.2007.12.002]

43.   Lloyd A, Anne Kennan M, Thompson KM, Qayyum A. Connecting with new information landscapes: information literacy practices of refugees. J Doc 2013 Jan 11;69(1):121-144. [doi: 10.1108/00220411311295351]

44. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]

45. Hersberger JA. Are the economically poor information poor? Does the digital divide affect the homeless and access to information? Can J Inf Libr Sci 2013 Oct 22;27(3):45-63. [doi: 10.29173/cais522]

46. Chatman EA. The impoverished life-world of outsiders. J Am Soc Inf Sci 1996 Mar;47(3):193-206. [doi: 10.1002/(sici)1097-4571(199603)47:3<193::aid-asi3>3.0.co;2-t]

47. Britz JJ. To know or not to know: a moral reflection on information poverty. J Inf Sci 2016 Jul;30(3):192-204. [doi: 10.1177/0165551504044666]

48. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]

49. Palumbo R, Adinolfi P, Annarumma C, Musella M. The special information needs of low health literate patients. Exploratory insights from an Italian survey. In: Proceedings Management in a Digital World. Italia: Decisions, Production, Communication; 2016 Presented at: XXVIII Sinergie Annual Conference Referred Electronic Conference; 2016; Udine, Italia.

50. Suárez A, Rodrigo MJ, Muneton M. M. Parental activities seeking online parenting support: is there a digital skill divide? Revista de Cercetare si Interventie Sociala 2016;54:e36. [doi: 10.33788/rcis]

51. Pluye P, El Sherif R, Granikov V, Hong QN, Vedel I, Galvao MC, et al. Health outcomes of online consumer health information: a systematic mixed studies review with framework synthesis. J Assoc Inf Sci Technol 2019 Jul;70(7):643-659 [FREE Full text] [doi: 10.1002/asi.24178] [Medline: 31423458]

52. Gregor S. The nature of theory in information systems. MIS Q 2006;30(3):611. [doi: 10.2307/25148742]

53. Saracevic T. Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. J Am Soc Inf Sci 2007 Nov;58(13):1915-1933. [doi: 10.1002/asi.20682]

54. Leonardi. When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. MIS Q 2011;35(1):147. [doi: 10.2307/23043493]

55. Bush PL, Pluye P, Loignon C, Granikov V, Wright MT, Pelletier J, et al. Organizational participatory research: a systematic mixed studies review exposing its extra benefits and the key factors associated with them. Implement Sci 2017 Oct 10;12(1):119 [FREE Full text] [doi: 10.1186/s13012-017-0648-y] [Medline: 29017557]

56. Bush PL, Pluye P, Loignon C, Granikov V, Wright MT, Repchinsky C, et al. A systematic mixed studies review on Organizational Participatory Research: towards operational guidance. BMC Health Serv Res 2018 Dec 22;18(1):992 [FREE Full text] [doi: 10.1186/s12913-018-3775-5] [Medline: 30577859]

57. Creswell J, Plano Clarke V. Designing and conducting mixed methods research. Thousand Oaks: SAGE; 2018.

58. Pluye P, Bengoechea EG, Granikov V, Kaur N, Tang DL. A world of possibilities in mixed methods: review of the combinations of strategies used to integrate qualitative and quantitative phases, results and data. Int J Mult Res Approaches 2018 Jul 16;10(1):41-56. [doi: 10.29034/ijmra.v10n1a3]

59. Roelen K, Camfield L. Mixed methods research in poverty and vulnerability: sharing ideas and learning lessons. London: Palgrave Macmillan; 2015.

60. Cappelleri J, Zou K, Bushmakin A. Patient-reported Outcomes: Measurement, Implementation and Interpretation. New York, USA: CRC Press; 2014.

61. Granikov V, Grad R, El Sherif R, Shulha M, Chaput G, Doray G, et al. The Information Assessment Method: Over 15 years of research evaluating the value of health information. EFI 2020 Apr 03;36(1):7-18. [doi: 10.3233/efi-190348]

62. Pluye P, Granikov V, Bartlett G, Grad RM, Tang DL, Johnson-Lafleur J, et al. Development and content validation of the information assessment method for patients and consumers. JMIR Res Protoc 2014 Feb 18;3(1):e7 [FREE Full text] [doi: 10.2196/resprot.2908] [Medline: 24550180]

63. Sherif RE, Roy P, Tang DL, Doray G, Dubois M, Bush P, et al. The value of user feedback: Parent's comments to online health and well‐being information providers. In: Proc. Assoc. Info. Sci. Tech. 2017 Oct 24 Presented at: ASIST; Oct 2017; Washington DC p. 662-663. [doi: 10.1002/pra2.2017.14505401107]

64. Bujold M, El Sherif R, Bush PL, Johnson-Lafleur J, Doray G, Pluye P. Ecological content validation of the Information Assessment Method for parents (IAM-parent): A mixed methods study. Eval Program Plann 2018 Feb;66:79-88 [FREE Full text] [doi: 10.1016/j.evalprogplan.2017.09.011] [Medline: 29053984]

65. Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment 1995 Sep;7(3):238-247. [doi: 10.1037/1040-3590.7.3.238]

66. Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess 2004 Sep;16(3):231-243. [doi: 10.1037/1040-3590.16.3.231] [Medline: 15456379]

67. Lowry R. The confidence interval for the difference between two independent proportions web calculator. 2019. URL: http://vassarstats.net/prop2_ind.html [accessed 2020-10-05]

68.   Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998 Apr 30;17(8):873-890. [doi: 10.1002/(sici)1097-0258(19980430)17:8<873::aid-sim779>3.0.co;2-i] [Medline: 9595617]

69.   Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]

70.   Bazeley P, Jackson K. Qualitative data analysis with NVivo. Thousand Oaks: SAGE; 2013.

71.   Fereday J, Muir-Cochrane E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. International Journal of Qualitative Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]

72.   Weber M. Économie et société: les catégories de la sociologie. Paris: Plon; 1995.

73.   Pampalon R, Hamel D, Gamache P, Simpson A, Philibert MD. Validation of a deprivation index for public health: a complex exercise illustrated by the Quebec index. Chronic Dis Inj Can 2014 Feb;34(1):12-22 [FREE Full text] [Medline: 24618377]

74.   Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The Relationship of Health Literacy With Use of Digital Technology for Health Information: Implications for Public Health Practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]

75.   Keller J, McDade K. Attitudes of low-income parents toward seeking help with parenting: implications for practice. Child Welfare 2000;79(3):285-312. [Medline: 10813085]

76.   MacPhee D, Fritz J, Miller-Heyl J. Ethnic Variations in Personal Social Networks and Parenting. Child Development 1996 Dec;67(6):3278. [doi: 10.2307/1131779]

77.   Henly JR, Danziger SK, Offer S. The contribution of social support to the material well-being of low-income families. J Marriage and Family 2005 Feb;67(1):122-140. [doi: 10.1111/j.0022-2445.2005.00010.x]

78.   Malone M, While A, Roberts J. Parental health information seeking and re-exploration of the 'digital divide'. Prim Health Care Res Dev 2014 Apr;15(2):202-212. [doi: 10.1017/S1463423613000194] [Medline: 23676618]

79.   Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]

80.   Kotchick B, Forehand R. Putting parenting in perspective: a discussion of the contextual factors that shape parenting practices. J Child Fam Stud 2002;11(3):225-269. [doi: 10.1023/A:1016863921662]

81.   Belsky J. The determinants of parenting: a process model. Child Dev 1984 Feb;55(1):83-96. [doi: 10.1111/j.1467-8624.1984.tb00275.x] [Medline: 6705636]

82.   Jones TL, Prinz RJ. Potential roles of parental self-efficacy in parent and child adjustment: a review. Clin Psychol Rev 2005 May;25(3):341-363. [doi: 10.1016/j.cpr.2004.12.004] [Medline: 15792853]

83.   Montigny F, Lacharité C. Perceived parental efficacy: concept analysis. J Adv Nurs 2005 Feb;49(4):387-396. [doi: 10.1111/j.1365-2648.2004.03302.x] [Medline: 15701153]

84.   Gerteis M. Through the patient's eyes: understanding and promoting patient-centered care. San Francisco: Jossey-Bass; 1993.

85.   Niela-Vilén H, Axelin A, Salanterä S, Melender H. Internet-based peer support for parents: a systematic integrative review. Int J Nurs Stud 2014 Nov;51(11):1524-1537. [doi: 10.1016/j.ijnurstu.2014.06.009] [Medline: 24998788]

86.   Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a Transactional Model of eHealth Literacy: Concept Analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]

87.   Reichwein B, Wolmarans L, Nantayi L, Nassali F, Kakinda A, Musumba D, et al. SegWeigh: a mixed-method approach to segmenting potential contraceptive user groups and meeting Family Planning 2020 goals. Int J Gynaecol Obstet 2015 Aug;130 Suppl 3:E8-E14 [FREE Full text] [doi: 10.1016/j.ijgo.2015.03.014] [Medline: 26024768]

88.   Luhmann N. What is Communication? Commun Theory 1992 Aug;2(3):251-259. [doi: 10.1111/j.1468-2885.1992.tb00042.x]

89.   Luhmann N. Social systems. Stanford: Stanford University Press; 1995:0804726256.

90.   Luhmann N. La confiance: un mécanisme de réduction de la complexité sociale. Paris: Economica; 2006.

91.   Luhmann N. La légitimation par la procédure. Québec: Les Presses de l'Université Laval; 2001.

92.   Vervier L, Valdez A, Ziefle M. "Should I trust or should I go?" or what makes health-related websites appear trustworthy? - an empirical approach of perceived credibility of digital health information and the impact of user diversity. In: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health. 2018 Presented at: 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health; March 2018; Madeira. [doi: 10.5220/0006734401690177]

93.   Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health Literacy and Use and Trust in Health Information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]

94.   Dutton WH, Shepherd A. Trust in the Internet as an experience technology. Information, Communication & Society 2006 Aug;9(4):433-451. [doi: 10.1080/13691180600858606]

95.   O'Neill O. Responses. In: Archard D, Deveaux M, Manson N N, Weinstock D, editors. Reading Onora O'Neill. London: Routledge; 2013:219-243.

96.   Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of Online Health Information: A Meta-Narrative Systematic Review. Am J Med Qual 2018;33(5):487-492. [doi: 10.1177/1062860617751639] [Medline: 29345143]

97.   Scheerder A, van Deursen A, van Dijk J. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics and Informatics 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]

98.   Ford-Gilboe M, Wathen CN, Varcoe C, Herbert C, Jackson BE, Lavoie JG, et al. How Equity-Oriented Health Care Affects Health: Key Mechanisms and Implications for Primary Health Care Practice and Policy. Milbank Q 2018 Dec;96(4):635-671 [FREE Full text] [doi: 10.1111/1468-0009.12349] [Medline: 30350420]

99.   Kutzin J. Health financing for universal coverage and health system performance: concepts and implications for policy. Bull World Health Organ 2013 Aug 01;91(8):602-611 [FREE Full text] [doi: 10.2471/BLT.12.113985] [Medline: 23940408]

100.   Lehoux P, Roncarolo F, Silva HP, Boivin A, Denis J, Hébert R. What Health System Challenges Should Responsible Innovation in Health Address? Insights From an International Scoping Review. Int J Health Policy Manag 2019 Feb 01;8(2):63-75 [FREE Full text] [doi: 10.15171/ijhpm.2018.110] [Medline: 30980619]

Abbreviations

IAM: Information Assessment Method
IP: internet protocol
N&G: Naître et grandir
SES: socioeconomic status

Review

The State of Evidence in Patient Portals: Umbrella Review

Marcy G Antonio[1], BSc, MPH; Olga Petrovskaya[2], RN, PhD; Francis Lau[1], PhD, FCAHS, CTSS

[1]University of Victoria, Victoria, BC, Canada
[2]University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Marcy G Antonio, BSc, MPH
University of Victoria
PO Box 1700 STN CSC
Victoria, BC, V8W 2Y2
Canada
Phone: 1 7783501089
Email: mantonio@uvic.ca

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/8/e32421

Abstract

Background: Patient portals have emerged as a recognized digital health strategy. To date, research on patient portals has grown rapidly. However, there has been limited evaluation of the growing body of evidence on portal availability, use, clinical or health behavior and outcomes, and portal adoption over time.

Objective: This paper aims to comprehensively consolidate the current state of evidence on patient portals using the umbrella review methodology, introduce our approach for evaluating evidence for quantitative and qualitative findings presented in included systematic reviews, and present a knowledge translation tool that can be used to inform all stages of patient portal adoption.

Methods: For this study, a modified version of the Joanna Briggs Institute umbrella review method was used. Multiple databases were searched for systematic reviews focused on patient portals, and the final sample included 14 reviews. We conducted a meta-level synthesis of findings from quantitative, qualitative, and mixed methods primary studies reported in systematic reviews. We organized the umbrella review findings according to the Clinical Adoption Meta-Model (CAMM). Vote-counting, GRADE (Grading of Recommendations, Assessment, Development, and Evaluations), and CERQual (Confidence in the Evidence from Review of Qualitative Research) were used to assess the umbrella review evidence.

Results: Our knowledge translation tool summarizes the findings in the form of an evidence map. Organized by the CAMM categories, the map describes the following factors that influence portal adoption and effects over time: patient contexts, patient's interest and satisfaction, portal design, facilitators and barriers, providers' attitudes, service utilization, behavioral effects, clinical outcomes, and patient-reported outcomes. The map lists the theories and mechanisms recognized in the included portal research while identifying the need for business models and organizational theories that can inform all stages of portal adoption. Our GRADE and CERQual umbrella review evaluation resulted in the majority of evidence being rated as moderate to low, which reflects methodological issues in portal research, insufficient number of studies, or mixed results in specific focus areas. The 2 findings with a high rating of evidence were patients' interest in using portals for communication and the importance of a simple display of information in the portals. Over 40 portal features were identified in the umbrella review, with communication through secure messaging and appointment booking mentioned in all systematic reviews.

Conclusions: Our umbrella review provides a meta-level synthesis to make sense of the evidence on patient portals from published systematic reviews. Unsystematic and variable reporting of portal features undermines the ability to evaluate and compare portal effects and overlooks the specific context of portal use. Research designs sensitive to the social, organizational, policy, and temporal dimensions are needed to better understand the underlying mechanisms and context that leverage the identified factors to improve portal use and effects.

(J Med Internet Res 2020;22(11):e23851) doi:10.2196/23851

XSL•FO

RenderX

Introduction

Background

Internationally, there has been an increasing effort to engage patients and consumers in their own health care using information and communication technology. The COVID-19 pandemic significantly stimulated the adoption and use of information and communication technology in primary care and outpatient clinics to facilitate remote visits, distant monitoring, and communication during the period of social distancing, particularly for patients living with chronic conditions [1-4]. These current events will further motivate various stakeholders to revisit the importance of eHealth tools, including electronic patient portals. A recent example from Canada is the province of Alberta that launched 2 patient portals in 2019, with a plan to gradually expand functionalities and patient engagement [5,6]. Since April 2020, patients and providers tested for COVID-19 across Alberta are able to access their test results online via MyHealth Records [5].

Countries such as Canada, the United Kingdom, and the United States have created national consumer digital health strategies and programs to encourage greater patient and consumer interactions with their health care providers through a variety of digital health solutions such as the patient portal [7-9]. For example, Infoway, an organization promoting Canada's health strategy, has funded several portal implementation projects over the last decade and produced benefit evaluation reports [8].

Patient portal is a digital health tool managed by a health organization to provide patients with secure online access to their own health information such as laboratory results, doctor's notes, and medication lists; care services such as appointment booking and reminders; and communication with their health care providers via secure messaging from anywhere via the internet [10]. It is also known as a tethered personal health record, which is a web-based interface linked to an electronic health record (EHR) where patients can view and interact with their health care data and providers [11].

Comparison With Prior Work

Despite widely acknowledged portal benefits, the empirical evidence on patient portals reflects important challenges and context-dependent outcomes of portal use. For example, a recent review examining the behavioral and clinical outcomes associated with portal use reported improved patient understanding of their health conditions and medication adherence while noticing modest or no effects on diabetes and hypertension biophysiological indicators [12]. Our previous review demonstrated that portal technology may inadvertently create health equity concerns by not paying sufficient attention to the social determinants of health during portal implementation [13].

Research on patient portals, including primary studies (PSs), systematic reviews (SRs), and meta-level reviews, has been rapidly growing. Most SRs focus on *specific* health conditions [14], patient populations [15], aspects of portal use and its effects or impact [16-18], or on a *select* study design, for example, randomized controlled trials (RCTs) using a portal as an intervention [16]. To date, 2 meta-level reviews have been published that summarize the findings on patient portals reported in SR papers [19,20]. van Mens et al [19] used the Clinical Adoption Framework to "map relationships between *determinant and outcome category*" from 19 SRs retrieved up to early 2018. Among the limitations of their review, van Mens et al [19] named inadvertent inclusion of duplicate PSs and the impossibility of evaluating the strength of evidence. The other review of 20 SRs summarized the methods (ie, study design), types of effects, and benefits of digital health interventions up to 2016 [20].

In addition to focusing on portal technology, both meta-level reviews [19,20] and several SRs [21-23] reported combined findings on a variety of eHealth tools, which may limit the reader's ability to discern portal-specific effects. Further, SRs may not always explicitly assess the quality of included PSs, thus potentially giving equal weight to findings characterized by various degrees of empirical support. When quality appraisals are included, the focus is often on RCTs and not broadened to other study designs [12,24]. Further, to the best of our knowledge, no published reviews summarizing evidence on patient portals have evaluated the strength of generated evidence as the concluding step of their review process. The limited quality assessments of the included studies and the absence of evaluations of the strength of outcomes or evidence offer readers little guidance on interpreting blanket statements of mixed and inconclusive results. To sum up, the various types of portal reviews that have been conducted to date present some substantive and methodological limitations mentioned earlier, thus providing an opportunity for our umbrella review.

Goal of This Study

Our contributions are threefold. The first is substantive: this umbrella review comprehensively consolidates the current state of evidence about patient portals. The second is methodological: we included a wide range of high-quality SRs that were specifically focused on portal technology, eliminated duplicate PSs, and appraised the quality of umbrella review quantitative and qualitative evidence using modified GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) and CERQual (Confidence in the Evidence from Review of Qualitative Research) criteria, thus presenting the relative strength of each umbrella review finding. The third contribution is knowledge translation, where our findings in the form of an evidence map can provide guidance for organizations in their patient portal adoption efforts. This is especially important in the Canadian context as many jurisdictions are actively pursuing patient portals at this time.

Methods

Objectives and Questions

The objectives of this umbrella review are to summarize the current state of evidence in patient portals based on published SRs and to create an evidence-based knowledge translation tool for the adoption of this technology. The questions addressed in this umbrella review are as follows:

1. What are the characteristics of the patient portals being introduced and used in different settings?
2. What are the system-related, health care provider–related, and patient-related factors that influence the introduction, use, and impact of patient portals?
3. What is the impact of patient portals on clinical outcomes of care?

Methodology

Our methodology is detailed in a published protocol [25] registered and updated in PROSPERO (International Prospective Register of Systematic Reviews; PROSPERO registration number CRD42018096657). We employed the Joanna Briggs Institute (JBI) umbrella review method [26] with *modifications* [25]. Overall, our umbrella review attempted to adhere to best practice methodological recommendations outlined by Pollock et al [27] and Smith et al [28].

Search Strategy and Inclusion or Exclusion Criteria

In April 2018, the original search was conducted in 9 databases: MEDLINE, EMBASE, CINAHL Plus with Full Text, Web of Science Core Collection, Scopus, the Cochrane Database of Systematic Reviews, the PROSPERO registry, the JBI Database of Systematic Reviews and Implementation Reports, and Proquest Dissertations and Theses. An updated search in MEDLINE was conducted in November 2019 to identify the SRs published since the initial search. The complete search strategy is provided in Multimedia Appendix 1. The key

inclusion criterion was specific to SRs focused on patient portals (irrespective of population groups and study designs) and published since 1990 in English. The SRs that were excluded were those with multiple eHealth technologies, standalone (ie, not tethered) personal health records, those focused on low- and medium-resource countries and thus contextually unique, reviews of reviews, scoping and integrative literature reviews, and reviews that do not provide a complete list of included PSs or designs.

Review Selection and Critical Appraisal

Citations were imported to Covidence. Two researchers independently screened titles and abstracts and then full-text articles against the inclusion and exclusion criteria. Discrepancies were resolved by consensus between 2 researchers and/or by a third researcher. The methodological quality of each SR was independently assessed by at least two researchers using the JBI critical appraisal checklist for SRs consisting of 11 questions [26]. Upon reaching a consensus among all researchers, low-quality SRs with a cutoff point <6 were eliminated.

Our initial database search for SRs yielded 158 citations. After eliminating duplicates and screening for relevant titles and abstracts, we retained 40 SRs for full-text assessment. By applying the inclusion or exclusion criteria, we identified 16 SRs for critical appraisal. An updated search in November 2019 identified 108 citations, yielding 6 SRs for full-text assessment, 3 of which underwent critical appraisal (the list of excluded reviews from the full-text assessment is provided in Multimedia Appendix 2 [14-18,24,26,29-36]). The 19 SRs included in critical appraisal had scores that ranged from 3 to 11, with 6 reviews scoring 10 or 11 (Multimedia Appendix 2). Five SRs with scores <6 were excluded, leaving 14 reviews as the final selection for synthesis [14-18,24,29-36]. Figure 1 shows the Preferred Reporting Items for SRs and Meta-Analyses flow diagram for the selection process.

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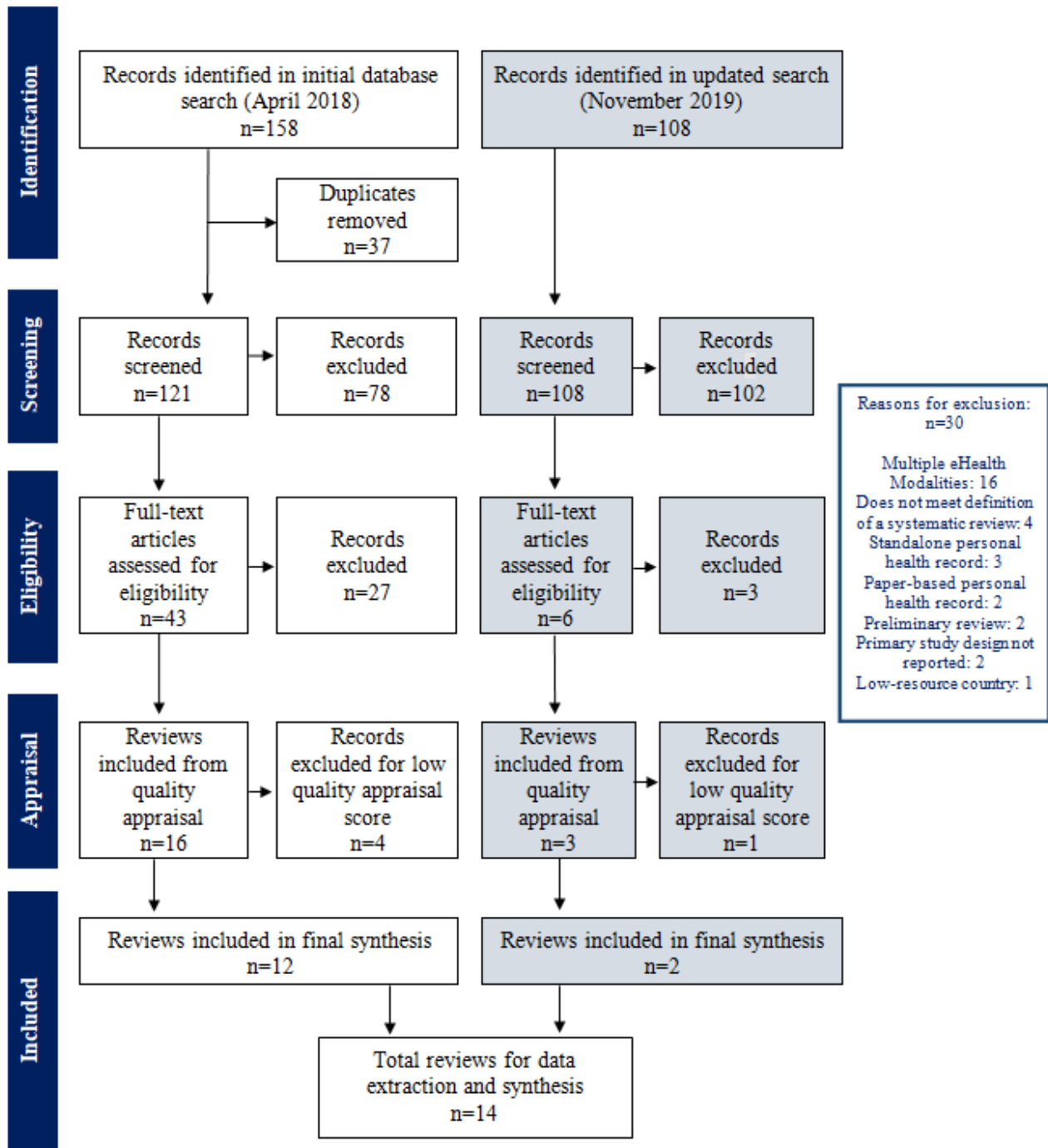
Analysis and Synthesis

In this paper, we refer to *the findings* at 3 different levels. PS findings are the individual findings reported by the SR authors in the results section and appendices of the SRs. In turn, SR findings represent the synthesis by SR authors reflecting how they combined PSs. The umbrella review findings are our synthesis of the PS findings reported in included SRs and of any conceptualizations or models advanced as the outcome in the included SRs.

Initially, we categorized the included SRs according to the logic underpinning their approach to synthesis [25,37]: reviews following the aggregation logic (n=13) in which the SR authors reported outcomes based on the summary of individual PS findings and reviews following the configuration logic (n=1) where the SR authors' analysis combined findings from PSs to articulate a theory. Figure 2 shows our categorization of the included SRs for data extraction and synthesis. We grouped all included *aggregation* SRs into (1) group A: purely quantitative reviews (n=3) where the SR authors only reported on findings from quantitative PSs, and (2) group B: reviews with a mix of quantitative, qualitative, or mixed methods PSs (n=10). There were no purely qualitative reviews (ie, those with exclusively qualitative PS design).

Figure 2. Data Extraction flow chart.
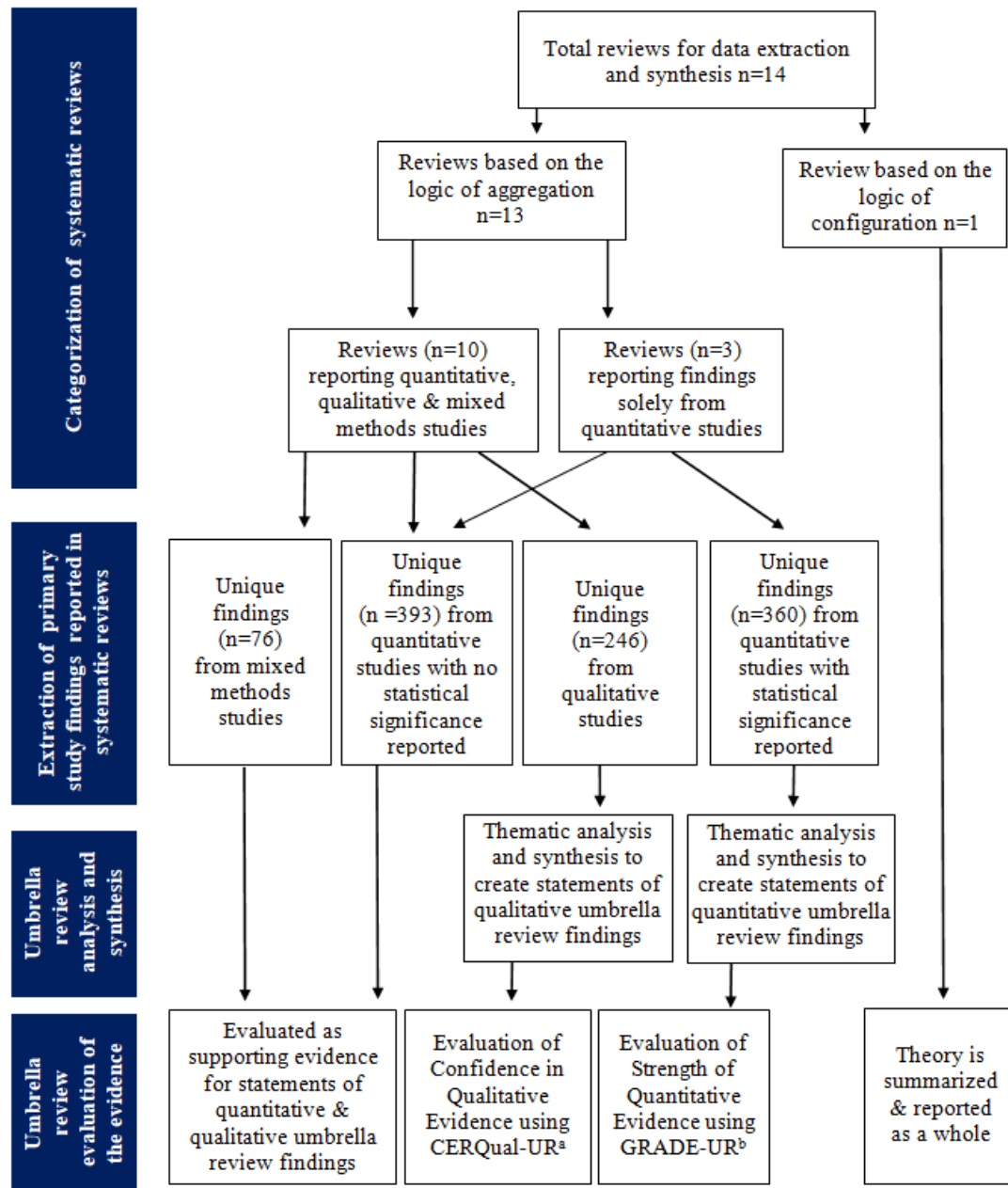
^a CERQual-UR: Confidence in the Evidence from Review of Qualitative Research at the Level of an Umbrella Review.

^b GRADE-UR: Grading of Recommendations, Assessment, Development, and Evaluations at the Level of an Umbrella Review.

One researcher extracted all the data, and 100% of the outputs were validated by at least one other researcher, with discrepancies resolved by discussions among the 3 researchers. All relevant PSs, their design, sample size, and all findings, as reported by the SR authors were extracted into Excel tables. Duplicate findings from duplicate PSs were removed to manage an overlap among the reviews. Data were synthesized into themes and then a smaller number of domains and statements. No statistical meta-analyses or subgroup analyses were performed due to heterogeneity of the study design.

As a precursor for rating the strength of umbrella review quantitative evidence and the confidence in umbrella review qualitative evidence, our 2 main data extraction documents were quantitative and qualitative Excel tables, respectively. The

headings and/or introductory sentences in the paragraphs describing findings or results in each SR were used for initial domain categorization. Domains were further developed through an iterative process of cross-checking with the Clinical Adoption Meta-Model (CAMM) framework, sorting within Excel documents, and weekly analysis meetings among the researchers.

Our quantitative data extraction table included those PSs from SRs (both group A and group B), where the SR authors referenced statistical significance (*P* values, significance, and confidence intervals) of PS findings. Our qualitative data extraction table included those PSs from SRs in group B, where the SR authors provided sufficient information about the study design for the reader to classify those PSs as qualitative. After

removing duplicates, the quantitative table included 360 unique individual findings from 10 SRs [14-17,30-35], and the qualitative table included 246 unique findings from 10 SRs [14,15,17,24,29-31,33,35,36]. Three SRs that were part of our umbrella review included papers that were a mix of PSs and reviews [17,24,35]. As an additional strategy to manage duplicates, we only extracted data from the PSs included in these SRs.

In addition, to extract all other data not fitting the 2 aforementioned tables, we created mixed methods and quantitative descriptive Excel data extraction documents, similar to the designs of PSs as reported in SRs. Our mixed methods table included 76 unique findings from mixed methods PSs from SRs in group B. The quantitative descriptive table included 393 unique findings from quantitative PSs from SRs in both groups A and B for which no statistical data or significance were provided by the SR authors, and which were unsuitable for evaluating the strength of evidence.

Rating the Umbrella Review Evidence

As the concluding step, 2 researchers independently assessed the strength of evidence for quantitative umbrella review finding statements and the confidence in the evidence for qualitative umbrella review finding statements. For this purpose, we developed meta-level umbrella review tools, GRADE-UR (Grading of Strength of Evidence for Quantitative Research at the Level of an Umbrella Review) and CERQual-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review), by applying a voting-counting method [38] and adapting GRADE [39-41] and CERQual [42-44] SR evaluation tools. The methodological approach was conceptualized independent from the GRADE working group. The GRADE-UR and CERQual-UR acronym was created by the authors of this paper to reflect an adaptation of GRADE and CERQual.

The GRADE-UR tool rates the strength of evidence as high, moderate, low, or insufficient [39-41]. Briefly, GRADE-UR evaluation of the quantitative finding statements is based on the following information: the SR authors' critical appraisal of PSs, the PS sample sizes, the number of RCTs, reporting of statistical significance, the number of PS findings that agree with the umbrella review finding statement, and the outcome measures used in PSs. The CERQual-UR tool rates confidence in the evidence as high, moderate, low, or very low. CERQual-UR evaluation of the qualitative finding statements is based on the following information: specific questions from our critical appraisal using the JBI checklist, SR methodological features, and their presentation of the results. A further description of GRADE-UR and CERQual-UR can be found in Multimedia Appendix 3 [26,38-44].

To the best of our knowledge, no tools are currently available for rating the evidence in reviews that synthesize mixed methods and findings from quantitative studies reported in SRs without any reference to statistical analysis; thus, we were unable to assess umbrella review evidence synthesized from our respective Excel tables. However, we compared umbrella review findings from mixed methods or quantitative descriptive PSs with the umbrella review findings from quantitative and qualitative PSs.

We looked for differences and correspondences with the goal of formulating any additional umbrella review finding statements arising from mixed methods or quantitative descriptive findings and not subject to GRADE-UR or CERQual-UR evaluation. This information can be found in Multimedia Appendix 4 [14-18,24,29-36].

Output of the Umbrella Review

The output of our umbrella review consists of 2 summary of findings and evidence profile tables, a narrative synthesis, and a knowledge translation tool in the form of an evidence map on patient portal use and impact. CAMM [45] underpins the categorization of these outputs. CAMM is a maturity model used to understand, describe, and explain the adoption of digital health technologies through the preadoption, early adoption, and mature adoption stages [45]. It is a temporal *adoption* model with 5 dimensions: availability, system use, clinical/health behavior, outcomes, and time. In this review, adoption refers to the planning, implementation, utilization, and support of a patient portal. *Availability* refers to making the patient portal accessible to users. *Use* refers to patterns of user interaction and experience with the portal. *Clinical/health behavior* refers to changes in user behaviors from interacting with the portal. *Outcomes* refers to the health impact from portal use. *Time* refers to the transition periods across the 4 dimensions [45].

Results

Review Characteristics

The population, objectives, design, and context of the 14 reviews are described below and further summarized in Multimedia Appendix 2. All 14 reviews were published between November 2012 and November 2019, 11 of which were published in 2015 or later. With duplicates removed, 280 unique PSs were identified across the 13 *aggregation* reviews, resulting in 1075 unique findings. We handled the findings of the only *configuration* review [18] holistically, without disintegrating into separate components. This realist review [18] proposed mechanisms for achieving portal outcomes.

Population

A total of 6 reviews focused on specific populations: people with diabetes [14,35] or other chronic conditions [17] and pediatric [15], vulnerable [32], and hospitalized [33] patients. Study participants included patients, family members, parents or guardians, and health care providers. A total of 7 reviews reported sample sizes of their included PSs, with sample sizes ranging from 5 participants in a qualitative study [33] to 529,605 in a cross-sectional study [31].

Objectives

Included SRs examined patient engagement [29], facilitators and barriers of portals use [36], meaningful use [17], health literacy [30], mechanisms for achieving portal outcomes [18], effects [17,24,31], and the impact of portals [16,35]. The findings from quantitative PSs reported in SRs pertained to portal enrollment or use levels by sociodemographic factors, the role of portal training, patient satisfaction and empowerment, clinical outcomes including screening rates and treatment

adherence, and portal impact on health service utilization [14-17,29,31,32,34]. The findings from qualitative PSs reported in SRs tended to focus on barriers and facilitators; portal design; communication between providers and patients; perceived care quality; providers' concerns; and patient safety, empowerment, engagement, and satisfaction [14,15,17,24,29-31,35,36].

Study Design

Across the 13 *aggregation* reviews, the number of PSs ranged from 5 [16] to 143 [24], yielding 280 unique PSs of varied designs. The quantitative PSs were classified as 32 RCTs, 18 cohort studies, 8 time series and 16 cross-sectional studies, 3 surveys, 8 pre-post, 2 post only, 1 retro-audit, and 1 quasi-experimental (further information on the design of the PSs for each of the UR findings is provided in Multimedia Appendix 4). In these quantitative studies, data were generated through structured self-report questionnaires, system logs, administrative data sets, and patient medical records. None of the reviews included meta-analyses due to the presence of heterogeneous population groups and diverse measures reported. The designs or methods of data collection used in qualitative and mixed methods PSs encompassed interviews, questionnaires, focus groups, usability studies including observations, and case reports. Critical appraisal tools were used in 7 reviews [14-16,24,31,32,34,35]. Heterogeneity was explicitly discussed as an issue in 4 reviews [16,24,32,34]. Limitations recognized by SR authors included the variability in portal utilization measures [14], portal types and population [31], and portal definitions [33], low quality of reporting [35], limited variety of study designs [31,33], and the potential of underpowered studies [32].

Across the 13 *aggregation* reviews, we found limited application of models, theories, and frameworks as the conceptual foundation for evaluating patient portal use and outcomes. None of the included SRs applied models, theories, or frameworks as a guide for their review. Some authors briefly referenced models, theories, and frameworks as justification for the review or in their discussion, whereas others summarized the use of models and theories in the included PSs [29,31,36]. Goldzweig et al [31] found that 5 out of 21 PSs used a theory or model, and Irizarry et al [29] found that 11 of 120 PSs used a theoretical framework, but the SR authors did not elaborate on how these theoretical underpinnings related to the outcomes. One SR observed that the chronic care model was cited most often within PSs [29], and 2 SRs commented on how the model's concept of self-management influences outcomes [14,24]. The following theories were referenced across the SRs: Roger's Diffusion of Innovation [29], activation theory [36], theory of coping and self-determination [36], and grounded theory [sic] [36]. Review authors recommended reconsidering the business model of patient care [34] and developing a framework to identify appropriate outcome measures for long-term portal use [36].

Context

A total of 10 reviews named the countries where the PSs were conducted; most were in the United States, with Europe, Australia, and Asia referenced [14-16,18,24,30,31,34-36]. A total of 8 reviews reported study settings, which varied from hospitals, clinics, group practices, and primary care, or a combination of settings [15,16,24,29,31,33,34,36].

Characteristics of Patient Portals

Portal features were reported inconsistently and unsystematically across reviews (a summary of the portal features described in the reviews is provided in Multimedia Appendix 2). Two SRs itemized and compared specific patient portal features reported across each PS [31,33]. Other SRs summarized portal features in their results section or appendices as an overview of the intervention, a collective summary sentence, or in reference to individual findings. Some reviews did not report specific portal features in their findings, but described portal functions in general, in the introductory section.

In total, we identified 41 portal features, with a range across reviews from 3 [30] to 25 [31]. Secure messaging or communication and appointment booking were the 2 features mentioned in all reviews. Other common features that we identified were access to laboratory and test results, visit summaries, and medication renewals. In contrast, portal features that allowed patients to generate data through care plans, patient self-assessment tools, journals, and the ability to edit data were mentioned less frequently. Frequently used features in a pediatric portal included immunization records, secure messaging, and appointment scheduling [15]. The SR by Kelly et al [33] on inpatient portals highlighted patients' desire to be able to view their daily schedule and information on medication dose, frequency, timing, administration, route, and side effects.

Seven reviews mentioned features that patients viewed as desirable but not commonly offered. These included proxy access, medication glossaries with photos, medication side effects and instructions, care goals or plans with feedback, symptom tracking, videoconferencing, portal access through onsite kiosks, voice recognition for older adults, and text messaging for quality assurance service [15,18,24,29-31,33]. Specific features desired by patients in an inpatient setting included hospital room number; health care provider names or photos; medical information on condition and what will happen next; recovery goals; and access to physician notes, operative reports, and test results [33].

Summary of Umbrella Review Findings, Evidence Profile, and the Knowledge Translation Tool

Tables 1 and 2 display the umbrella review finding statements and the evidence profile. To understand how the strength of the evidence and the confidence in the evidence were evaluated, refer to Multimedia Appendix 3. We found few examples of high confidence in the evidence and no examples of high strength of evidence. High confidence means that additional studies are unlikely to generate new findings on account of the topic being relatively well researched, SRs being of high quality, and the findings representing the phenomenon accurately. Moderate strength of evidence indicates that the finding is likely but there are some deficiencies in the current evidence [41]. Moderate confidence in the evidence indicates that the findings reasonably represent the phenomenon [43]. Low to very low confidence indicates that the topic is under-researched, studies have methodological weaknesses or inconsistent findings, and

new studies are likely to generate useful findings that can contradict existing evidence. Our umbrella review findings should be interpreted in this context, with the amount and quality of existing research represented in the included SRs and the consistency of findings playing an important role in the rating of evidence.

Our evidence map for patient portal adoption is presented in Figure 3. The map is a visual knowledge translation tool of the current state of evidence of patient portal use and impact. In this figure, the image of the 4 CAMM stages is from Price et al [45] and the 5 columns provided have been organized by the domain, and outcome or findings listed in Tables 1 and 2. In the remaining sections of our results, we present the umbrella review findings through the 4 stages of the map, while drawing attention to select strength of evidence and confidence in the evidence ratings.

Table 1. Summary of quantitative umbrella review findings and Grading of Strength of Evidence for Quantitative Research at the Level of an Umbrella Review evaluation of quantitative evidence.

Umbrella review domain and summary of quantitative findings statement (SR[a] source)	Strength of the evidence according to the GRADE-UR[b] criteria[c]
Patient characteristics	
Patients with better controlled diabetes are more likely to enroll or use a portal as compared to other patients with diabetes [14,31,35].	Moderate
Patients with private insurance in the US context are more likely to enroll or use a portal [14,15,31,34,35].	Moderate
Patients with higher illness(es) burden or need are more likely to enroll or use a portal [14,15,31,34].	Moderate
White people are more likely to enroll or use a portal [14,15,31,32,34].	Moderate
Middle-aged people (≤65 years) are more likely to enroll or use a portal [14,17,31,32,34].	Moderate
People who have a higher income are more likely to enroll or use a portal [14,15,31,35].	Moderate
Males with diabetes are more likely to enroll or use portal as compared with females with diabetes [14,35].	Moderate
Patients with higher health literacy are more likely to enroll or use a portal [14,17,30].	Low
Females are more likely to access online information and use a portal [17,31,34].	Low
People who have a higher education level are more likely to enroll in and use a patient portal [14,31,35].	Low
Patient-related facilitators	
Patients are more likely to register and use a portal after portal-related education and training [32].	Moderate
Patient satisfaction	
Patients who use patient portals report higher satisfaction with communication, treatment, medications, and care [16,31,34,35].	Moderate
Behavioral effects	
Use of patient portals can increase adherence, mostly medication adherence across different patient populations [16,17,31,33,34].	Moderate
Use of patient portals can improve screening, vaccinations, examinations, and/or care across different patient populations [31,34,35].	Moderate
Use of patient portals can improve visit preparation and communication and information sharing between patients and providers [14,16,31,34].	Low
Service utilization effects	
Health care provider's workload related to contacts and messaging does not change with patient portal adoption [34].	Moderate
Patients' access to social support and mental health and testing services does not change with portal use [31,33].	Moderate
Hospitalization rates do not change with patient portal use [16,31,34].	Low
Emergency department visits do not change with patient portal use [16,31,34].	Low
Phone or messaging volume received by health care providers does not change with patient portal use [16,17,31,34].	Low
Patient portal use results in an increase in office, primary care, specialist, outpatient, or after-hour visits [15-17,31,34,35].	Low
Patient portal use does not reduce no-show rates [17,34].	Low
Clinical outcomes	
There is improvement in HbA$_{1c}$[d] levels for patients with diabetes who use patient portals [15-17,31].	Moderate
There is improvement in LDL[e], HDL[f], cholesterol, or lipids for patients with diabetes who use patient portals [15,16,31,35].	Low
There is no change in systolic and diastolic blood pressure for patients with diabetes or hypertension who use patient portals [16,31,35].	Low
Psychosocial, cognitive function, BMI, symptom stability, and depression and anxiety status does not change across multiple patient populations who use patient portals [16,17,31].	Low

Umbrella review domain and summary of quantitative findings statement (SR[a] source)	Strength of the evidence according to the GRADE-UR[b] criteria[c]
Patient-oriented outcomes	
Patient empowerment and self-efficacy scores do not change with portal use [16,31].	Low

[a]SR: systematic review.

[b]GRADE-UR: Grading of Recommendations, Assessment, Development, and Evaluations at the Level of an Umbrella Review.

[c]Indicates the strength of the evidence and was calculated based on study limitations, directness, consistency, precision, and reporting of bias. The ratings are from high, moderate, and low. Any statements we evaluated as insufficient were moved to the supporting evidence tables in Multimedia Appendix 4.

[d]HbA$_{1c}$: hemoglobin A$_{1c}$.

[e]LDL: low-density lipoprotein.

[f]HDL: high-density lipoprotein.

Table 2. Summary of qualitative umbrella review findings and Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review evaluation of qualitative evidence.

Umbrella review domain and summary of qualitative findings statement (SR[a] source)	Confidence in the evidence according to the CERQual-UR[b] criteria[c]
Patients' interest in the potential of portals	
Patients are interested and satisfied in using patient portals if they are easy to use and useful [15,24,33].	High
Patients are interested in using patient portals for communication and opportunity to message providers [17,30,33].	High
Portal design and features	
Patients value information in patient portals that is easy to understand, written in lay or nonmedical language, transparent, and presented in a simple display [29,33].	High
Patients want prescription refills, and hospitalized patients in particular want information on medication that includes dose, frequency, timing, administration, route, and side effects [29,33].	High
Minimal navigation steps and educational information on specific laboratory results, medications, and allergies are important health equity and patient-friendly considerations [15,29,33].	Moderate
The information within patient portals gives patients and parents a greater sense of control, involvement, understanding, and security in care planning [15,33,35].	Moderate
Patients appreciate the scheduling function in patient portals, such as booking appointments online and scheduling, and daily planning in inpatient setting [15,29,33].	Low
System-related factors	
Guideline development, framework for governance, and compliance with regulations are important for integrating patient portals into organizational processes [24,33].	Moderate
Patient-related facilitators	
Use of patient portals is facilitated by the enhanced communication over traditional methods and positive patient-provider interactions and relationships [14,33,36].	Low
Encouragement and instruction on patient portals offered by providers and families is a facilitator of portal use [14,29,36].	Low
Patient-related barriers	
Patient barriers to portal use and enrollment include time, limited system knowledge, lack of awareness of patient portals and related features, and doubt or lack of belief in portal benefits or value [14,17,29,36].	Moderate
Technical barriers to portal use and enrollment include type of interface, lack of technical or computer skills or training or support or literacy, lack of computer or internet access, and forgotten passwords [14,17,24,30,31,33,35,36].	Moderate
Unauthorized access, privacy, security, and trust or confidentiality concerns are barriers to portal use and enrollment [14,15,24,29,30,33,36].	Moderate
Patients' lack of desire in enrolling and using portals relates to their preferences and satisfaction with existing means of communication [14,17].	Very low
Providers' attitudes and concerns	
Providers are concerned about liability and increases or changes in workload, and the lack of training, skills, and resources for using patient portals and prefer to have support staff screen messages [24,29,33].	Moderate
Providers are concerned that the information contained in portals may overwhelm, cognitively overload, or increase patients' anxiety and that patient-generated data may be inaccurate [24,29,33].	Moderate
Providers perceive patient portals could encourage patient engagement, and secure messaging could support communication of complex information, while having concerns about impact on patient-provider relationships [24,29,33].	Low
Providers are concerned about patient safety, privacy, and confidentiality and prefer control over access and authentication of users to protect the information in patient portals [24,33].	Low
Lack of incentive and reimbursement may result in providers being less engaged with portals than patients may assume and instructing patients not to use [14,31].	Low
Usability-related barriers	

Umbrella review domain and summary of qualitative findings statement (SR[a] source)	Confidence in the evidence according to the CERQual-UR[b] criteria[c]
Usability-related barriers which result in negative experiences and use of patient portals include: reminders and messages that are unreliable, have a slow response, or may not directly reach providers, and information that is inaccurate or difficult to locate due to complex navigation, visual layout, and language [14,29-31].	Low
Patient satisfaction	
Online communication with providers outside their hours is preferred by patients and parents, as it is easier to understand, more convenient, supports accessing test results, and allows for timely and consistent responses [15,29,33,35].	Moderate
Patient safety	
Patient portals enhance efficiency and patient safety when patients find and request correction of errors, especially medication errors [17,24,33,35].	Moderate
Patients with limited health and computer literacy value portal use, but safe and effective use may be compromised by an inability to interpret results and having to take longer to complete patient portal tasks [29-31].	Low
Behavioral effects	
Patient portals can facilitate access to medical information that can engage and empower patients to be confident in their self-management and current care [29,31,36].	Low
Service utilization effects	
Patient portals can impact provider workload by increasing number of phone calls or emails or secure messaging and length of face-to-face visits [17,24,35].	Very low
Patient-oriented outcomes	
Patient portals empower patients in shared decision making, prepare for visits, enable better expression of ideas and concerns, and encourage engagement in self-care and self-management [17,24,35].	Moderate
Patient portals support communication, enhance discussions, and shift power relations between patients and providers [17,24,29,33,35].	Moderate
Patient portals can improve quality of care and caregiver experience and reduce care burden [17,33].	Low

[a]SR: systematic review.

[b]CERQual-UR: Confidence in the Evidence from Review of Qualitative Research at the Level of an Umbrella Review.

[c]Indicates the confidence in the evidence and was calculated based on methodological limitations, coherence, relevance, and adequacy. Ratings are from high, moderate, low, and very low.

XSL•FO
RenderX

Figure 3. Evidence Map Across Portal Adoption Stages.

Who Adopts Patient Portals

Many SRs examined the characteristics of patients who use patient portals. Portal users are more likely to be middle aged (≤65 years) [14,17,31,32,34]; White (compared to Black, Hispanic and Asian) [14,15,31,32,34]; and have higher income [14,15,31,35], private insurance [14,15,31,34,35], higher education [14,31,35], or higher illnesses burden or service needs [14,15,31,34]. Females are more active adopters of patient portals when compared to males [17,31,34], except for those with lower socioeconomic status [34]. Among diabetes patients,

males are more likely to enroll or use portals [14,35]. Patients with lower health literacy and numeracy skills are less likely to use portals [14,17,30].

Patients' Interest and Satisfaction in Portal Design

Patients are interested and satisfied in using patient portals if they are designed to be easy to use and useful [15,24,33]. The information in patient portals should be transparent and easy to understand, written in nonmedical language, and have a simple display [29,33]. To make portals more patient-friendly and attuned to variations in health literacy, there should be minimal

navigation steps and educational information on specific laboratory results, medications, and allergies [15,29,33].

Factors to Consider in Increasing Portal Adoption by Patients

There are many barriers that impede the adoption of patient portals. Patients may forget their password, not have internet access, or lack the necessary interface or literacy and computer skills [14,17,24,30,31,33,35,36]. Portal use may also be discouraged when patients receive unreliable messages, encounter complex language, or have difficulty locating information within the portal [14,29-31]. Patients may also be unaware about portals, or have limited belief in portal benefits [14,17,29,36] and have privacy and security concerns [14,15,24,29,30,33,36]. Facilitation of portal adoption included encouragement from providers and family members [14,29,36] and training and education on portals [32].

System-Related Factors for Portal Implementation

A limited number of reviews addressed these factors. The lack of incentive or reimbursement may dissuade providers from promoting portal use to their patients [14,31]. Two SRs referenced the need for the development of guidelines and governance frameworks for integration of patient portals into organizational processes and to support compliance with regulations [24,33].

Health Care Providers' Concerns on Portal Implementation

Multiple qualitative studies reported on providers' attitudes toward portals [24,29,33]. The accuracy of patient-entered information and how portal information can increase patient anxiety or overwhelm, overload, and offend patients were repeated concerns of providers [24,29,33]. Providers preferred control over information to offset concerns about patient safety, privacy, and confidentiality [24,33]. Providers were also concerned about liability; change in workload; and the lack of training, skills, and resources for using patient portals [24,29,33].

Impact on Communication and Patient-Provider Relationships

Communication and influence on patient-provider relationships became a common thread throughout our evidence map. Some SRs found that patients were interested in the potential of using portals for communication with providers [17,30,33] and preferred the convenience, asynchronous aspect, and timeliness of communication afforded within patient portals [15,29,33,35]. Patients who used portals reported satisfaction with communication and the resulting treatment, medications, and care supported through portals [16,31,34,35]. However, some patients lack interest in portals, as they are satisfied with their current mode of communication [14,17].

Three SRs reported on providers' interest in portals for encouraging patient engagement and communication of complex information with recognition of the potential impact on patient-provider relationships [24,29,33]. In addition, the evidence that portal use improved communication, information sharing, and patient-provider relationships [14,16,31,34] was rated as low.

Portal Use and Impact on Patient Outcomes and Behavioral Effects

Although patient portals may improve patient safety by having patients note and correct errors [17,24,33,35], safe and effective portal use may be compromised because of patients' limited health and computer literacy [29-31]. Portal use may also reduce caregiver burden and improve the quality of preventive or follow-up care [17,33]. For clinical outcomes, evidence is limited and inconsistent. We found low strength of evidence for changes in blood pressure [16,31,35] and metabolic measures [15,16,31,35] with portal use. We found moderate strength of evidence for improvement in hemoglobin A_{1c} [15-17,31], preventive care [31,34,35], and medication adherence [16,17,31,33,34].

Impact on Service Utilization and Provider Workload

Limited number of studies have examined changes in the utilization of preventive or testing services [31,33], rate of hospitalizations [16,31,34] and emergency department visits [16,31,34], and no-show rates [17,34], all of which demonstrated low strength of evidence. Although a significant number of quantitative and qualitative PSs examined portal impact on provider workload, the evidence was inconsistent as to how patient portals may impact the number of contacts and face-to-face consults, phone volume, emails or messaging, and office visits [15-17,24,31,34,35]. We found very low confidence in the evidence that portals change provider workload and moderate strength of evidence for no change in workload. A similar pattern of variability was found within quantitative descriptive PSs not included in the aforementioned description: 10 found no change, 12 reported a decrease, and 20 reported an increase in provider workload.

Mechanisms

The review by Otte-Trojel et al [18] included in our analysis hypothesized 4 mechanisms for patient portals to achieve outcomes: patient insight, convenience, continuity of care, and patient activation [18]. Insights that patients gain from their online health information and EHR can improve communication, empowerment, understanding of one's health condition, and adherence to treatment. Convenience is the time saved when patients have online access to providers and services. Care continuity improves patient-provider communication. Activation leads to empowerment through power balance and self-identity and to better self-care through improved relationships, trust, and availability of educational resources [18]. When we examined how similar concepts were referenced in the other included SRs, we found the following: low strength of evidence for association between portal use and changes to patient empowerment and self-efficacy scores [16,31]; moderate confidence in the evidence that portals could empower patients for self-care and shared decision making [17,24,35]; and moderate confidence in the evidence that portal information could provide patients and parents a greater sense of control, involvement, understanding, and security in care planning [15,33,35].

Time Dimension in Portal Use

The study by Grossman et al [32] was the only SR that reported in detail how portal use changed over time as a result of portal training and education. Bush et al [15] observed that longitudinal studies did not track portal usage over time. Other SRs did not seem to extract information on changes in portal use over time. Similarly, longitudinal changes in clinical outcomes were not highly represented in SRs, with 2 reviews referencing changes in outcomes over multiple time periods [31,35].

Additional Findings From Mixed Methods or Quantitative Descriptive PSs

Data included in our mixed methods and quantitative descriptive extraction tables and not subject to the evidence rating provided additional support for the quantitative and qualitative umbrella review findings discussed earlier (Multimedia Appendix 4). In addition, noteworthy findings from this group of studies that are not found in our evidence tables include providers preferring emails focused on simple, self-limiting problems [24]; the importance of tailored messages sent to patients [29] and proxy access (family member access and caregiver access to information in the portal) [15,30,31,33]; and lower uptake of patient portals than initially anticipated [15,24]. Of note, proxy access and portal uptake are addressed only in a few PSs.

Discussion

Principal Findings

To the best of our knowledge, our umbrella review demonstrates the first attempt to adapt GRADE and CERQual processes and to develop GRADE-UR and CERQual-UR tools for evaluating evidence generated in an umbrella review. Moreover, our combined approach of evaluating the evidence and application of CAMM at the umbrella review level provides a novel approach for analyzing outcomes. In particular, we demonstrate how this approach can provide a more nuanced understanding of the evidence for common findings generated in portal research.

Our umbrella review provides an evidence map based on CAMM and the consolidated summary on the current state of evidence on patient portals and can be used to inform all stages of portal adoption. The map should be used in conjunction with the evidence tables to understand the strength of the available empirical support for different factors influencing patient portal adoption, use, and outcomes. In the next four paragraphs, we present key umbrella review findings and elaborate on how the map can be applied to address current knowledge gaps and across research, industry, policy, and practice.

The temporal aspect of CAMM suggests that portal adoption per se is not a guarantee of its effectiveness and that it should not be evaluated at a single point in time; rather, there are transitions from *interest* to *registration or enrollment*, *activation*, and then to *use or utilization* ending with *empowerment* in the best-case scenario. In our umbrella review, we found enrollment and use often being conflated, thus blurring the line between these 2 separate, yet very significant, dimensions of portal adoption.

We color coded different factors included in the evidence map to visually represent the strength of the current evidence. The colors can reveal not only existing knowledge gaps but also methodological assumptions made in some research on portals. For example, the evidence for the statement *portals change provider workload* was rated as very low, indicating that further research on this topic is likely to produce useful findings. However, the tables summarizing our findings also show that this topic has been extensively researched, but the studies exhibited great variability in the direction of the findings. This incongruence in the findings about changes to provider workload reveals a complex interplay of factors mobilized when technology is introduced into clinical practice and patients' homes. Neither the types of studies that control for these factors (eg, RCTs) nor the study designs that simply explore participants' satisfaction and perception are capable of shedding light on local organizational contexts that are often responsible for divergent portal outcomes.

The ideal use of the evidence map would be that each stakeholder looks across adoption stages and recognizes the interdependency among different factors, as the following examples suggest (italics indicate factors within the map). For example, for industry to support the promised long-term vision of improving clinical and patient-oriented outcomes, an *easy-to-use portal design* is required that does not introduce *usability barriers* at a later stage. In the policy realm, there is a need for *governance guidelines* that address both patients' and providers' *privacy and security concerns* to encourage trusting *patient-provider relationships*.

In addition, the evidence map reveals the areas of differing values that need to be considered within practice to achieve successful portal adoption. For example, patients are *interested in the communication* afforded by patient portals, whereas providers are concerned about the *increased workload* introduced by the new channels for communication. Further, portals are promoted as a tool for *patient empowerment*, whereas providers are concerned about *the loss of control* over information. With portals designed to serve more than one user group and to support their sometimes-divergent agendas, the success of portal technology hinges on acknowledging these different values and addressing these differences through the engagement of all relevant user groups.

Limitations

There are a number of strengths to this umbrella review. First, the elimination of duplicate PSs from the included SRs provides a more accurate account of the reported findings. Second, the application of vote-counting, GRADE-UR, and CERQual-UR allows the direction, strength, and confidence of the evidence to be quantified and compared. Third, the evidence-based knowledge translation tool offers practical guidance to those involved in the planning, implementation, and support of patient portals. The evidence map seems promising as it helps to cast the use and impact of patient portals over time across preadoption, early adoption, and mature adoption stages while summarizing both key known success factors strongly supported by research and areas with low evidence base where more research is needed.

There are possible limitations to our search strategy. First, the literature on patient portals is varied in its coverage of eHealth modalities, inconsistent in defining portals, and evolving. Our search and selection strategy could have missed reviews that should be included. Second, we only focused on reviews published in English; those in other languages could have been missed. We also relied on the SR authors' identification of patient portals. However, in reviewing the SRs' reference list, we discovered that some SRs had included PSs that were broader in scope than the SR definition of patient portals. We accounted for these occasional discrepancies between the intended and actual SR focus during our GRADE-UR and CERQual-UR evaluation.

Our application of GRADE and CERQual at the meta-synthesis level could have been flawed, as there is little or no guidance available on how to appraise evidence synthesized across several SRs. In particular, we used SR authors' reporting of statistical significance and *P* values to evaluate the strength of evidence. Many of the PSs included in the GRADE-UR evaluation had a limited sample size, and reporting on the power calculation of these studies was lacking; therefore, we were unable to evaluate these claims of statistical significance.

Similarly, we relied on SR authors' reporting of the PS findings. When we suspected possible errors in SR authors' reporting of findings, we consulted the original PS; in these cases, consensus among all umbrella review researchers was sought to determine how to address these individual discrepancies. However, this process was not exhaustive, and we did not cross-check all the findings between SRs and PSs.

Comparison With Prior Work and Suggestions for Future Research

At the time of analyzing the included SRs, we noticed a certain tendency for not recognizing the findings from qualitative research as evidence worthy of being evaluated for its strength, which may reinforce the hierarchy of what is considered evidence. This observation might be explained by the lack of appropriate tools such as CERQual, which was developed recently and presents a counterpart to GRADE. The notion of *evidence* in the context of portal research should not be limited to a narrowly conceived evidence stemming from controlled studies. The role of RCTs is well recognized, and the strength of this design is acknowledged [20]; however, the findings and issues raised in our umbrella review and other reviews [46] call for a broader conceptualization of evidence. There are promising emergent qualitative designs that focus on the social, organizational, political, policy, and local context through emergent, ethnographic, and co-design approaches [47,48].

When examining portal-related outcomes, comparative studies to date have focused on biomedical measures for people with diabetes or hypertension and patient empowerment scores for general patient populations. We recommend further research with people living with other chronic conditions and comparative measures that communicate patients' values and perspectives [49]. This will involve greater integration of patient-oriented measures that can evaluate the outcomes that are of greatest interest to patients and experiences with their portal-enabled health care encounters.

Unsystematic reporting of portal features in the included SRs may reveal a mistaken assumption in some patient portal research that portal features are inconsequential for the outcomes. In comparative research, this can lead to a faulty comparison of portals that may be quite different from each other. We recommend that researchers report actual portal functionalities and technology characteristics. Ammenwerth et al [50] offer a simple yet useful portal taxonomy that includes 7 functionalities (access, remind, request, communicate, share, manage, and educate), which can be easily used when describing portal features to support comparisons across settings and studies.

Consistent with the findings of other SRs not included in our umbrella review, we found that the most extensive areas of research and evidence were on patient-related factors, namely, common barriers and facilitators [22,51], clinical and behavioral outcomes [20], and the role of patient demographic factors in portal adoption [19,20,22,46,51]. Health care provider–related factors were primarily focused on provider concerns [22,46,51], with a lack of examples on how these concerns can be addressed. SR reporting of a lack of guidelines and business models to guide portal implementation was the only statement we identified under system-related factors. Our umbrella review extends the understanding of these patient-, provider-, and system-related factors by reporting and rating the evidence at a meta-level. In addition, we identified current evidence gaps related to proxy access, portal uptake, and most significant to our review, the need for theoretical frameworks sensitive to system-level factors. We aimed to address this theoretical gap in applying CAMM to the development of our evidence map for patient portal adoption.

Conclusions

Our umbrella review offers an organized knowledge translation tool on what is known about patient portals, the quality of the available evidence, and areas that require further work. The evidence map can be used to inform planning, implementing, and supporting the adoption of patient portals across research, industry, policy, and practice. Through our GRADE-UR and CERQual-UR approach, we demonstrated not only how to consolidate findings from SRs, including PSs of various designs, but also how to evaluate the strength and confidence in the evidence of findings from quantitative and qualitative studies. For many of the umbrella review findings, the quality of the evidence was rated as low. This suggests at least two interrelated conclusions. For many identified factors playing a role in portal success, their interactions and underlying mechanisms, especially over time, are still mostly unknown and invite new research. Along with this, we need a broader conceptualization as to what constitutes evidence. This calls for study designs and theoretical perspectives attentive to the contextual complexity of portal adoption.

Acknowledgments

Authors' Contributions

MA and FL developed the intellectual idea for the review. OP and FL drafted the protocol and its various components, with MA commenting on drafts. OP led the submission of the protocol and PROSPERO registration. During early stages, FL led the development of the study, with MA and OP providing suggestions. OP collaborated with a librarian to confirm the original search strategy and managed citations in Covidence. MA and OP updated the search. FL, MA, and OP independently conducted critical appraisal. FL extracted most of the data, with MA and OP validating and/or completing data extraction. FL synthesized quantitative findings, with MA independently validating and/or completing the synthesis. FL led vote-counting and application of GRADE, with MA independently validating and/or completing the evaluation of quantitative evidence. FL drafted the early versions of the manuscript and the knowledge translation tool. MA led all subsequent steps of this study. MA synthesized qualitative, mixed methods, and quantitative descriptive findings, with OP independently validating the synthesis. MA led the adaptation of CERQual tool, development of CERQual-UR method, and synthesis of qualitative statements, with OP contributing throughout the process. MA and OP independently applied CERQual-UR to evaluate the confidence in qualitative evidence. MA developed the final knowledge translation tool, with OP providing feedback. MA led the manuscript development and authored several parts, with OP writing specific parts and contributing throughout. MA, OP, and FL approved the final manuscript and helped to resolve disagreement and reach consensus throughout the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 20 KB - jmir_v22i11e23851_app1.pdf]

Multimedia Appendix 2
Exhibits for extraction of patient portal umbrella review.
[XLSX File (Microsoft Excel File), 40 KB - jmir_v22i11e23851_app2.xlsx]

Multimedia Appendix 3
Application of GRADE-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review) and CERQual-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review): An example.
[PDF File (Adobe PDF File), 127 KB - jmir_v22i11e23851_app3.pdf]

Multimedia Appendix 4
Additional support for GRADE-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review) and CERQual-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review) findings.
[PDF File (Adobe PDF File), 271 KB - jmir_v22i11e23851_app4.pdf]

References

1. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for COVID-19. Br Med J 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
2. Vandekerckhove P, Vandekerckhove Y, Tavernier R, de Jaegher K, de Mul M. Leveraging user experience to improve video consultations in a cardiology practice during the COVID-19 pandemic: initial insights. J Med Internet Res 2020 Jun 25;22(6):e19771 [FREE Full text] [doi: 10.2196/19771] [Medline: 32519964]
3. Lindsay JA, Hogan JB, Ecker AH, Day SC, Chen P, Helm A. The importance of video visits in the time of COVID-19. J Rural Health 2020 Jun 7:- epub ahead of print [FREE Full text] [doi: 10.1111/jrh.12480] [Medline: 32506751]

XSL•FO
RenderX

4. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying connected in the COVID-19 pandemic: telehealth at the largest safety-net system in the United States. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]

5. Using MyHealth Records. Government of Alberta. -. URL: https://myhealth.alberta.ca/ myhealthrecords#:~:text=You%20need%20to%20sign%20up,License%20or%20Alberta%20Identification%20card [accessed 2020-05-01]

6. Killen V. Introducing: MyHealth Records Patient Portal!. Alberta Medical Association. 2019 Jan 17. URL: https://www. albertadoctors.org/leaders-partners/ehealth/myhealth-records/introducing-mhr-patient-portal [accessed 2020-05-01]

7. Online Health and Prescription Services. National Health Service. 2019 Oct 15. URL: https://www.nhs.uk/using-the-nhs/ nhs-services/gps/gp-online-services/ [accessed 2020-07-21]

8. Citizen Health Portal Improves Access and Helps Patients Better Manage Their Care. Canada Health Infoway. 2020. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/videos/ 3546-citizen-health-portal-improves-access-and-helps-patients-better-manage-their-care?Itemid=101 [accessed 2020-07-20]

9. Patient Access Information for Providers. HealthIT. 2019 Nov 6. URL: https://www.healthit.gov/topic/ patient-access-health-records/providers [accessed 2020-07-20]

10. What is a Patient Portal? HealthIT. 2017 Sep 29. URL: www.healthit.gov/faq/what-patient-portal [accessed 2020-08-01]

11. Are There Different Types of Personal Health Records (PHRs)? HealthIT. 2019 Jul 30. URL: https://www.healthit.gov/ faq/are-there-different-types-personal-health-records-phrs [accessed 2020-08-06]

12. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]

13. Antonio MG, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 1;26(8-9):871-883. [doi: 10.1093/jamia/ocz054] [Medline: 31066893]

14. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]

15. Bush RA, Connelly CD, Fuller M, Pérez A. Implementation of the integrated electronic patient portal in the pediatric population: a systematic review. Telemed J E Health 2016 Feb;22(2):144-152 [FREE Full text] [doi: 10.1089/tmj.2015.0033] [Medline: 26258289]

16. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]

17. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]

18. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]

19. van Mens HJ, Duijm RD, Nienhuis R, de Keizer NF, Cornet R. Determinants and outcomes of patient access to medical records: systematic review of systematic reviews. Int J Med Inform 2019 Sep;129:226-233. [doi: 10.1016/j.ijmedinf.2019.05.014] [Medline: 31445260]

20. Zanaboni P, Ngangue P, Mbemba GI, Schopf TR, Bergmo TS, Gagnon M. Methods to evaluate the effects of internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018 Jun 7;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]

21. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. The influence of patient portals on users' decision making is insufficiently investigated: a systematic methodological review. Int J Med Inform 2018 Mar;111:100-111 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.028] [Medline: 29425621]

22. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]

23. Andrikopoulou E, Scott P, Herrera H, Good A. What are the important design features of personal health records to improve medication adherence for patients with long-term conditions? A systematic literature review. BMJ Open 2019 Sep 26;9(9):e028628. [doi: 10.1136/bmjopen-2018-028628] [Medline: 31558449]

24. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep 8;4(9):e006021. [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]

25. Petrovskaya O, Lau F, Antonio M. Synthesising evidence on patient portals: a protocol for an umbrella review. BMJ Open 2019 Mar 30;9(3):e024469. [doi: 10.1136/bmjopen-2018-024469] [Medline: 30928933]

26. Aromataris E, Fernandez R, Godfrey C, Holly C, Khali H, Tungpunkom P. Umbrella reviews. In: Joanna Briggs Institute Reviewers' Manual. Adelaide, South Australia: The Joanna Briggs Institute; 2017.

27. Pollock A, Campbell P, Brunton G, Hunt H, Estcourt L. Selecting and implementing overview methods: implications from five exemplar overviews. Syst Rev 2017 Jul 18;6(1):145 [FREE Full text] [doi: 10.1186/s13643-017-0534-3] [Medline: 28720141]

28. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb 3;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]

29. Irizarry T, deVito AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

30. Coughlin SS, Stewart JL, Young L, Heboyan V, de Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]

31. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]

32. Grossman LV, Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 1;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]

33. Kelly MM, Coller RJ, Hoonakker PL. Inpatient portals for hospitalized patients and caregivers: a systematic review. J Hosp Med 2018 Jun 1;13(6):405-412 [FREE Full text] [doi: 10.12788/jhm.2894] [Medline: 29261819]

34. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151. [doi: 10.3399/bjgp15X683941] [Medline: 25733435]

35. Mold F, Raleigh M, Alharbi NS, de Lusignan S. The impact of patient online access to computerized medical records and services on type 2 diabetes: systematic review. J Med Internet Res 2018 Jul 6;20(7):e235 [FREE Full text] [doi: 10.2196/jmir.7858] [Medline: 29980499]

36. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]

37. Sandelowski M, Voils CI, Leeman J, Crandell JL. Mapping the mixed methods-mixed research synthesis terrain. J Mix Methods Res 2012 Oct;6(4):317-331 [FREE Full text] [doi: 10.1177/1558689811427913] [Medline: 23066379]

38. Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. J Am Med Inform Assoc 2010;17(6):637-645 [FREE Full text] [doi: 10.1136/jamia.2010.004838] [Medline: 20962125]

39. Berkman ND, Lohr KN, Ansari M, McDonagh M, Balk E, Whitlock E, et al. Grading the strength of a body of evidence when assessing health care interventions for the effective health care program of the agency for healthcare research and quality: an update. AHRQ Method Effect Health Care 2013 Nov 18;13(14):1-44. [Medline: 24404627]

40. Gibbons MC, Wilson RF, Samal L, Lehmann CU, Dickersin K, Lehmann HP, et al. Consumer health informatics: results of a systematic evidence review and evidence based recommendations. Transl Behav Med 2011 Mar;1(1):72-82 [FREE Full text] [doi: 10.1007/s13142-011-0016-4] [Medline: 24073033]

41. Berkman ND, Lohr KN, Ansari MT, Balk EM, Kane R, McDonagh M, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. J Clin Epidemiol 2015 Nov;68(11):1312-1324. [doi: 10.1016/j.jclinepi.2014.11.023] [Medline: 25721570]

42. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci 2018 Jan 25;13(Suppl 1):2 [FREE Full text] [doi: 10.1186/s13012-017-0688-3] [Medline: 29384079]

43. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 2015 Oct;12(10):e1001895 [FREE Full text] [doi: 10.1371/journal.pmed.1001895] [Medline: 26506244]

44. Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 2: how to make an overall CERQual assessment of confidence and create a summary of qualitative findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]

45. Price M, Lau F. The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decis Mak 2014 May 29;14:43 [FREE Full text] [doi: 10.1186/1472-6947-14-43] [Medline: 24884588]

46. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? A systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]

47. Greenhalgh T, Shaw S, Wherton J, Hughes G, Lynch J, A'Court C, et al. SCALS: a fourth-generation study of assisted living technologies in their organisational, social, political and policy context. BMJ Open 2016 Feb 15;6(2):e010208. [doi: 10.1136/bmjopen-2015-010208] [Medline: 26880671]

XSL•FO

RenderX

48.   Locock L, Robert G, Boaz A, Vougioukalou S, Shuldham C, Fielden J, et al. Using a national archive of patient experience narratives to promote local patient-centered quality improvement: an ethnographic process evaluation of 'accelerated' experience-based co-design. J Health Serv Res Policy 2014 Oct;19(4):200-207. [doi: 10.1177/1355819614531565] [Medline: 24840387]

49.   Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice--adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. J Clin Epidemiol 2013 Aug;66(8 Suppl):S12-S20 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.04.005] [Medline: 23849145]

50.   Ammenwerth E, Lannig S, Hörbst A, Muller G, Schnell-Inderst P. Adult patient access to electronic health records. Cochrane Database Syst Rev 2017;6:CD012707 [FREE Full text] [doi: 10.1002/14651858.CD012707]

51.   Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient portal use in diabetes management: literature review. JMIR Diabetes 2018 Nov 6;3(4):e11199 [FREE Full text] [doi: 10.2196/11199] [Medline: 30401665]

Abbreviations

CAMM: Clinical Adoption Meta-Model
CERQual: Confidence in the Evidence from Review of Qualitative Research
CERQual-UR: Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review
EHR: electronic health record
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
GRADE-UR: Grading of Strength of Evidence for Quantitative Research at the Level of an Umbrella Review
JBI: Joanna Briggs Institute
PS: primary study
RCT: randomized controlled trial
SR: systematic review

XSL•FO
RenderX

Original Paper

Role of Technology in Self-Assessment and Feedback Among Hospitalist Physicians: Semistructured Interviews and Thematic Analysis

Andrew Lukas Yin[1,2*], BA, MBA; Pargol Gheissari[2*], BEng, MS, MSIS; Inna Wanyin Lin[2*], BA, BSc, MEng; Michael Sobolev[2,3], PhD; John P Pollak[2,4], PhD; Curtis Cole[4,5], MD; Deborah Estrin[2,5], PhD

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

[2]Cornell Tech, New York, NY, United States

[3]Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States

[4]Department of Medicine, Weill Cornell Medicine, New York, NY, United States

[5]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Andrew Lukas Yin, BA, MBA
Medical College
Weill Cornell Medicine
1300 York Avenue
New York, NY,
United States
Phone: 1 212 746 5454
Email: aly2011@med.cornell.edu

Abstract

Background: Lifelong learning is embedded in the culture of medicine, but there are limited tools currently available for many clinicians, including hospitalists, to help improve their own practice. Although there are requirements for continuing medical education, resources for learning new clinical guidelines, and developing fields aimed at facilitating peer-to-peer feedback, there is a gap in the availability of tools that enable clinicians to learn based on their own patients and clinical decisions.

Objective: The aim of this study was to explore the technologies or modifications to existing systems that could be used to benefit hospitalist physicians in pursuing self-assessment and improvement by understanding physicians' current practices and their reactions to proposed possibilities.

Methods: Semistructured interviews were conducted in two separate stages with analysis performed after each stage. In the first stage, interviews (N=12) were conducted to understand the ways in which hospitalist physicians are currently gathering feedback and assessing their practice. A thematic analysis of these interviews informed the prototype used to elicit responses in the second stage.

Results: Clinicians actively look for feedback that they can apply to their practice, with the majority of the feedback obtained through self-assessment. The following three themes surrounding this aspect were identified in the first round of semistructured interviews: collaboration, self-reliance, and uncertainty, each with three related subthemes. Using a wireframe, the second round of interviews led to identifying the features that are currently challenging to use or could be made available with technology.

Conclusions: Based on each theme and subtheme, we provide targeted recommendations for use by relevant stakeholders such as institutions, clinicians, and technologists. Most hospitalist self-assessments occur on a rolling basis, specifically using data in electronic medical records as their primary source. Specific objective data points or subjective patient relationships lead clinicians to review their patient cases and to assess their own performance. However, current systems are not built for these analyses or for clinicians to perform self-assessment, making this a burdensome and incomplete process. Building a platform that focuses on providing and curating the information used for self-assessment could help physicians make more accurately informed changes to their own clinical practice and decision-making.

XSL•FO

RenderX

Introduction

As we explore a world where machine learning and advanced algorithms are increasingly used to assist in medical care, consideration of how to empower individuals to use these data themselves is often neglected. The aim of this project was to explore which tools or modifications to existing systems could most benefit hospitalist physicians, matching both their intrinsic and professional desires for self-assessment and improvement. Although applicable to many clinicians, hospitalist physicians represent a group that is specifically challenged in obtaining such feedback, as they do not see the patients they treat after discharge. Accordingly, they lack the opportunity to adjust or realize the benefits and shortcomings of the care provided. As a team of technologists and clinicians, we set out to understand how hospitalist physicians seek feedback to improve their practice, and to explore potential technical solutions that might be able to support them using a qualitative approach grounded in an iterative design process.

Embedded in the culture of medicine, physicians embrace the value of lifelong learning [1]. The pursuit of this goal is engrained in medical trainees, particularly in the form of self-directed learning and self-assessment [2], and is reinforced in the Physician Competency Reference Set produced by the American Association of Medical Colleges [3], the Common Program Requirements for residents by the Accreditation Council for Graduate Medical Education [4], and the requirements to participate in continuing medical education throughout a clinical career [5]. Nevertheless, it is recognized that self-assessment is a flawed, challenging process that in the wrong environment can adversely affect the participating clinicians. Physicians have a limited ability to accurately assess themselves, resulting in a feeling of imposter syndrome that can further reduce the physician's trust in other forms of self-assessment [6-8].

As a result, self-directed learning and self-assessment are commonly complemented by other models aimed to support the growth of the physician. One favored model is audit and feedback, a process that aims to show an individual whether or not performance is on par with a desirable target metric. This model has led to various behavior changes in physicians, including improving antimicrobial stewardship, completing relevant tasks such as discharge summaries, and helping inform appropriate use of tests and screens [9-12]. However, even with clear target metrics, the effectiveness of this method depends on baseline performance, how the feedback gets delivered, and the clarity of the action plan [13-15]. A second method is peer-to-peer or observation-based assessment, which ideally involves direct observation of the feedback receiver by the feedback giver, an appropriate institutional culture, and appropriate delivery of the feedback [16]. Thoroughly studied in medical training, peer-to-peer assessment has strengths in consistency and predictability of long-term quality, but greatly depends on trust, time, and attention to confidentiality [17-22]. Third, physicians can use information from patient feedback surveys to guide learning, but studies have shown that there are many contextual factors that influence actionable change. These include an appropriate work culture that embraces use of patient feedback, credibility of collection, and specific data content [23-25]. Other studies have unfortunately shown that physicians have at best a mixed view of patient feedback if they use it at all, with a negative view being more likely due to distrust of credibility, use to threaten their jobs, or distrust of the administrative motivations for using this feedback [26-28].

This study is based on the recognition that although there has been substantial research focusing on helping physicians to improve, progress is still needed to gain a deeper understanding about the feedback physicians currently use and the type of feedback they prefer [29,30]. In addition, although electronic medical record systems have been studied in the context of quality improvement [31,32], there appears to be a lack of tools or research described in the literature aimed at using existing technology and data available to support individual physician improvement.

Methods

Participants

Hospitalist medicine physicians from a single, large academic medical institution were invited to participate in this study. Randomly selected members in the department were emailed with information about the study to ask for their participation and were offered the option to speak in person, over the phone, or over video conferencing platforms. A total of 44 individuals were emailed with 17 responses obtained, 5 of whom were not interviewed due to scheduling limitations or loss in follow-up communications. Twelve individuals met our criteria and agreed to participate in the first round of interviews. A randomly selected subset of 4 participants were then interviewed again in the second round of interviews. All participants provided informed, written consent for their participation in the study and were not compensated for their participation. This study received an exempt status from the Cornell University Institutional Review Board under protocol number 1912009284.

Demographics

Among our interviewees, the average total years of work experience as a hospitalist was 5.8 years with the average time working at the current institution being 3.9 years. Five of the twelve participants were female. Ten of the participants described teaching as one of the main activities in their current role.

Data Collection: Semistructured Interviews Round 1

Semistructured interviews were used for a thorough examination of each participant's experience and motivations. Our interview methods closely mirrored those described previously [33,34]. Interviews were conducted in a private space either in person, over the phone, or over video conferencing platforms depending on the participant's preference and availability. Each interview lasted about 45 minutes and was composed of a single study

XSL·FO

RenderX

participant and two members from the research team. All interviews were conducted by AY, PG, and IL to provide consistency across interviews. One team member asked questions while the other took notes on any observable body language, tone of response, and potentially overlooked content. The interviewer used a set of predetermined open-ended questions developed following a previously described framework to understand all ways that participants pursue improvement in their work, seeking to understand how they get feedback, how they seek to improve, and what information they wish they could obtain [35]. This interview technique allowed for participants to dive deeply into their experiences and speak openly about them. All interviews were recorded and transcribed by members of the research team and combined with the notes taken during the interview itself. AY, PG, and IL reviewed each audio file and transcript for consistency and accuracy.

Data Analysis

A thematic analysis was performed following the methods of Braun and Clarke [36] to analyze the data owing to its strengths of accessibility and flexibility in analysis, enabling deep exploration of a rich dataset and to identify repeated concepts. The thematic analysis covered six phases: (1) familiarizing, transcribing, and reviewing all recordings and notes, and rereading each transcript prior to proceeding; (2) generating initial codes individually to allow for diversity in perspective and then discussing collaboratively to combine and rework codes; (3) searching for themes and subthemes that emerged as consistent patterns in coded elements; (4) reviewing potential themes and subthemes in relation to the dataset to refine themes and codes; (5) defining and naming themes and subthemes to best express the final concepts captured; and (6) producing a report [36]. After individual coding of two transcripts, the team developed a set of unique codes used for the remaining transcripts. AY, PG, and IL met to revise this codebook every two transcripts. Saturation, the point at which no new codes were needed, was reached after the 10th transcript. After completing the coding of all interview transcripts, the codes were consolidated into 22 categories by consensus, with final consolidation through group discussion leading to the final unifying themes and subthemes.

Wireframe Development

Using the identified themes and subthemes, the research team developed a wireframe of an app aimed at addressing user needs. Development of the wireframe involved an interdisciplinary team of technologists (DE, JP, PG, IL), clinicians (CC, AY), and a behavioral scientist (MS) to create a wireframe that addressed the identified user needs. This wireframe was then used for the second round of interviews.

Data Collection: Round 2 of Semistructured Interviews With the Wireframe

The research team set out to assess how well the wireframe addressed the themes and subthemes using semistructured interviews [33,34]. These interviews occurred with a subset of the original interview participants, allowing interviews to more quickly delve into more substantive topics. Each interview lasted about 30 minutes. The researchers showed each screen of the wireframe, using a script to give a basic overview followed by a set of predetermined open-ended questions [35]. Interviewees were shown the wireframes through video conference screen share or on one of the researcher's computers. All interviews were recorded and transcribed. Analysis focused on determining how the prototype addressed the themes and subthemes.

Results

Thematic Analysis of First Round Interviews

The research team identified themes and subthemes related to self-directed assessment and learning. The three themes that emerged were: collaboration, self-reliance, and uncertainty. Within each theme, three subthemes were extracted from the interviews with participants (Table 1).

Table 1. Descriptions of themes and subthemes.

Themes and subthemes	Description
Theme I: Collaboration	
I.I: Physicians are looking for feedback	Whether new or seasoned, physicians use many currently available resources to gather feedback for themselves but have difficulty sharing feedback with one another and are still looking for more for themselves.
I.II: Physicians have specific people they consult for feedback	In reviewing past decisions, physicians defer to a close ring of current or past colleagues or family members in the medical field rather than purposefully finding external experts or people they are less comfortable with.
I.III: Physicians interpret feedback more negatively than likely intended	Although feedback is usually not intended to do so, physicians tend to have a strong, negative emotional response to the feedback they receive.
Theme II: Self-Reliance	
II.I: Physicians have go-to resources for learning	Participants have an array of resources they readily rely on for learning, with little variation among participants.
II.II: Physicians build workarounds	Physicians build and maintain workarounds to obtain information of interest, but they find them to be inefficient and time-consuming.
II.III: Medicine can feel like a solo sport	Although the practice of medicine is commonly thought of as a team environment, participants sometimes feel they are acting on their own, which can be challenging.
Theme III: Uncertainty	
III.I: Physicians like numbers but need more context	With the current data available, physicians find it challenging to create actionable learning points as the data usually do not capture enough context.
III.II: Physicians cannot always find the "right" answer	There may not be a "right" answer in medicine much of the time, and physicians have a hard time finding out if their actions in the past were the best actions given the situation at the time.
III.III: Physicians' actions are limited by uncertainty about workplace cultural expectations	Uncertainty with workplace cultural or collaborative expectations creates friction for them in giving feedback.

Theme 1: Collaboration

Physicians rely on one another for feedback and point-of-care advice. With respect to feedback from colleagues, three subthemes arose. First, physicians do not receive much feedback in their daily work even when looking for it. Second, when they do solicit feedback, they typically only go to the people they trust. Last, on the occasions that physicians do receive feedback from one another, they tend to interpret the feedback more negatively than it is likely intended.

Subtheme 1.1: Physicians Are Looking for Feedback

Most of the respondents mentioned a lack of feedback in their daily work, a trend observed among both new and more experienced physicians. Some participants reflected on having more established learning and feedback mechanisms in residency. As one of the interviewees said: "there's not enough feedback in our careers, as soon as you finish medical school and residency and then you just go out into the world and, uh, basically get no feedback unless you look for it." [participant 3]

As a result, physicians are trying to get this feedback.

> *When I do my signout and email my signout to the next provider I will usually say if there is anything that is blatant or if you are noticing something that you want to comment on or give me feedback on then please do because there aren't many other ways for hospitalists to get feedback on their clinical judgment.* [participant 5]

When asked about giving feedback to other physicians, an interviewee mentioned that they have trouble giving direct constructive feedback: "I try to give some feedback very sneakily just by trying to give an update… by telling them how that [patient] is doing clinically, which I think implies some feedback." [participant 8]

Subtheme 1.2: Physicians Have Specific People They Consult for Feedback

Although a physician may ask for second opinions from experts or colleagues during real-time point of care, they tend to only go to people they trust when reflecting on whether a previous diagnosis could have been different. Interviewees mentioned these inner circles being friends from medical school, trusted colleagues, or family members in the medical field: "I definitely look to some of my most trusted colleagues to try to debrief and go over if I really did mess up, could have done something different, what would I do next time, what would you have done, etc." [participant 12]

Subtheme 1.3: Physicians Interpret Feedback More Negatively Than Likely Intended

On the occasions that physicians receive feedback, they share a tendency to feel negatively about that feedback. Although it is unlikely that the intention of the feedback giver was to create these negative emotions, participants have a notably negatively biased emotional response, even about items that were not necessarily their fault. One participant described this as:

XSL•FO
RenderX

I think most people, if you tell them "Hey, you could work on X, Y, Z things," and I see this when I talk to residents and students and have felt this myself in that position, that you feel a little personally attacked. [participant 3]

Others described this emotion as pervasive and strong. After hearing some feedback, one participant said:

I felt so awful, and the thing is I didn't even know the patient was in this situation. But, I think it made me feel so bad that it heightens your awareness after something that like...you're so scared and don't wanna repeat your mistakes so you overcompensate. [participant 6]

These feelings extend beyond negative feedback alone. Even in receiving positive feedback, participants expressed having a hard time feeling genuinely good if the feedback was not constructive or clear. One participant said:

It's not all positive feedback that I find useless but it's that very generic like obviously not very thoughtful feedback that I find useless. And then I can't even feel good about it because I don't even know what I did to deserve that comment. I feel like I don't even own that comment. [participant 5]

Theme 2: Self-Reliance

Participants shared that when it comes to taking actions and making decisions in challenging settings, they often rely on their own resources, skills, logic, and experience. Three main self-reliance trends were observed: using go-to resources, building workarounds, and managing the lonely nature of the work.

Subtheme 2.1: Physicians Have Go-To Resources for Learning

Almost all participants shared a list of literature resources that they heavily rely on when it comes to learning and decision-making, including UpToDate, journals such as *New England Journal of Medicine* and *Journal of the American Medical Association*, and even social platforms such as Twitter. One participant described:

I use UpToDate the most in my clinical work to help me work through an answer or read a little more of background about treatment... I use Twitter to passively scroll for new ideas or new papers, but I will save a tweet if I thought it was especially helpful that I could look at it later. [participant 12]

In addition, physicians are reviewing their past patients in the medical record, as one participant described:

After we get off service we tend to stalk the list and read about them and what our colleagues did. If there's a patient that I want to follow long-term you can make your own list on the electronic medical record and keep track of them that way. You can see what happens in the long term. [participant 6]

Subtheme 2.2: Physicians Build Workarounds

Although they have some resources, many participants stated that they create additional workarounds to find or keep relevant information. Some create lists of patients to track and revisit, while others keep documents with facts that they have learned from different sources. They expressed that these self-made systems take a lot of time and energy to maintain. One participant mentioned that:

I have an ad-hoc system which is not easy to do. Essentially, after completing every block, I compile a list of patients on that block, then I manually enter each patient ID into the electronic medical record. It's not easy to track patients, so I have to move them over, one by one. Then I have a set up where I can see their last admission date and their last outpatient date. I keep tabs on patients that are still active or that I want to keep an eye on...it's helpful to see what happens when they get discharged; was my judgment ok, did they come back. It's helpful to see how our doctors synthesize their problem, and if they get readmitted. That is normally a flag for me, I want to see whether they were readmitted because I could've done something differently, I want to know about and see what I can do in the future. Sometimes their disease progresses and I can't do anything about it, but it is also helpful to know. [participant 10]

Another participant described:

I have a long and incredibly messy running document on Evernote that's just quick pearls or facts that I've just learned along the way and I type into that every once in a while, either on my phone or on my computer. It's just saved there, and I don't do a very good job of organizing it or improving the layout of it. I will just search for a specific word and find it in the document, but it could certainly be much more elegant. [participant 8]

Although some participants have been able to continue their workarounds, some have given up along the way as things can quickly become too hard to keep up with or keep organized:

I never really looked at my ongoing notes again. It just took way too much time and it's something that I never looked at again. I think it points more towards organization; I need to be more organized in terms of where to store things so that I can retrieve it easily...I tried doing it on the computer with folders like a folder of evidence, physical diagnosis, or by disease. I have resources for cirrhosis, heart failure, and pneumonia and all these things, I have even aligned them with the competencies of hospital medicine. I even tried different things; I definitely need to be excellent in organizing my folders, but you just never go back to it. [participant 7]

Subtheme 2.3: Medicine Can Feel Like a Solo Sport

Although the practice of medicine is commonly thought of as a team environment, participants expressed that at times they feel a strong burden that they must do certain things on their

own—a burden that they can find challenging. Some noted this feeling in relation to the need to monitor their own performance. One participant noted that:

> *If a colleague takes over a patient and totally changes my plan and ends up being right about new management, there is no way for me to know about that without me looking. If I admit somebody, and I say that they have pneumonia and they ended up having a PE [pulmonary embolism], basically nobody tells me that I was wrong about that, which is insane.* [participant 3]

Another remarked:

> *When it comes to my clinical decision making, I think I have taken responsibility for feedback upon myself. When I go off service, for at least the next 2 or 3 days afterwards I pop into the charts of the patients that I've passed off and see what happened.* [participant 9]

Others noted a similar feeling related to staying current on advances in medical knowledge:

> *I think there are moments where I wonder if my level of knowledge is as up-to-date as it ought to be. I think there are areas that I probably could look up more out of curiosity, I just haven't.* [participant 1]

Theme 3: Uncertainty

Participants expressed that one of the most difficult aspects of assessing themselves is the uncertainty in their daily environment, making some actions and assessments challenging. This centered around three subthemes: uncertainty with data, with finding a correct answer, and with cultural and collaborative expectations.

Subtheme 3.1: Physicians Like Numbers but Need More Context

Almost all participants openly expressed the importance of data in driving their decisions and actions in real-time clinical work, but when it came to assessing their past performance, they expressed a lack of data that they trusted beyond rereading the clinical courses of patients. One participant explained: "the problem with feedback in hospital medicine, it's hard to get outcomes data or change of care data because so many hospitalists are in charge of someone." [participant 4]

In addition, participants expressed the challenge of using qualitative feedback from students and residents, as they commonly offer conflicting or unactionable comments:

> *they are usually very generic...so someone will say "she doesn't teach enough" and someone else will say "she teaches too much, rounds go on too long" and so I have noticed that oftentimes the constructive feedback only brings down my mood but doesn't really add much.* [participant 5]

When asked to discuss what data they would be interested in having, participants openly expressed interest in data about readmissions, length of stay, and mortality, but were quick to

mention reservations. One participant described the challenge with variance in the data:

> *I would love to know in general what of the patients I see are more or less likely to be readmitted compared to my peers or compared to some average, but the problem is, for example, 6 months into this job that I would imagine there would be so much variation just by pure random chance that it would probably take at least a sample size of 1 to 2 years to be able to tell us if any of our numbers were actually reflective of true differences from the mean or whether it was just statistical wackiness.* [participant 8]

Another described the challenge of figuring out how to attribute data to a single person:

> *It can get pretty noisy particularly at the physician level because physicians hand off a lot. One of the reasons the length of stay data is an issue is if I am on service for 7 days half of the people I discharge, I wasn't really involved with their care upfront.* [participant 2]

In other words, participants express difficulty finding reliable data that feels applicable, actionable, and timely. Even when discussing potentially valuable data points, participants worry that these data points do not consider the context and variability associated with each person's practice.

Subtheme 3.2: Physicians Cannot Always Find the "Right" Answer

Although they have doubts about some of the data points, participants are nevertheless looking to know if they made the right decisions. Currently, they primarily do this by checking in on a patient's record for some period of time to understand the outcome of the patient, even though this can still be challenging. One participant described:

> *If I look back at an old H & P [History and Physical], it may not reflect the correct diagnosis for this person because I didn't really get a handle on how to deal with the situation until day 2 or day 3. There isn't really a clear way as it's currently formulated to mark where it was in the chart that you had your breakthrough and figured out what it was and figured out what you wanted to do, so it requires a lot of wading through.* [participant 8]

Even with the ability to look back, it can be hard to find an exact answer. Participants expressed managing this in different ways, but many acknowledged how this uncertainty is a part of their job and challenging to navigate. One participant said:

> *If things aren't clear later on and we don't have the data to figure out whether they will be clear, we'll just have to move on. I think dealing with uncertainty is an important part of our job...it would be paralyzing to try to make sure that for every diagnosis we made or every treatment we offered, we had an exact knowledge of whether that was right or wrong. It's just impossible to know for some of those cases.* [participant 9]

Another described this challenge of parsing through what to assess and having too many factors to be able to clearly delineate what might be right or wrong as:

> *You could do something perfectly right and it's still difficult to know because a patient was so sick and had a bad outcome, there could be some emotional things tagged to it like if they're young or if you didn't get along with the patient and the family back in heightened emotions and make it difficult to. Or there is some conflict or you know you and a consultant weren't seeing eye to eye so there are a lot of different factors.* [participant 12]

As a result, participants expressed different ways of managing this challenge. One explored this a step further by stating the importance of reasoning as opposed to assessing whether or not the decision was correct, stating:

> *I guess rather than what the ultimate answer was, it matters more if I got there for the right reason and did the appropriate workup. Unless there was a right reason to making that decision, then it doesn't feel good to make the right decision for the wrong reason and to not have the right thought process behind it.* [participant 3]

Subtheme 3.3: Physicians' Actions Are Limited by Uncertainty About Workplace Cultural Expectations

Participants expressed that uncertainty with cultural or collaborative expectations creates friction for them in giving feedback. One remarked how normal conversations can become uncomfortable due to this issue, stating:

> *If there's not a cultural expectation, then it feels like a big deal even if you want it to not be a big deal and just be casual, quick feedback. It always makes the thing feel bigger than it is.* [participant 8]

Another commented on how this lack of structure can create challenges for the feedback giver and discourage future willingness to share, saying that: "There is no mechanism and my experiences have been quite negative in terms of actually being able to provide feedback to my colleagues." [participant 1]

Participants added how uncertainty with feedback structures create major stresses, even in important moments when they feel feedback to be necessary. One participant said: "It was scary because he's a senior attending and I was in my second year, but I felt that it was something that couldn't not be told." [participant 12]

In another way, this lack of expectations can cause a lot of stress as they try to parse through messages thinking there might be hidden feedback. One participant shared:

> *I'm sure the implicit message is here are the updates and here are the considerations that should've taken place. I'm sure that happens, but I don't think that's ever explicitly stated like "I would've done this rather than that." I can't read the mind of the person delivering updates, but I do wonder if that's what they're trying to convey just by human nature. Are*

> *they telling me this just because or is it subliminal messaging or am I just being delusional?* [participant 7]

One participant described the cultural challenge in balancing this:

> *I have done a lot of reading about the concept of uncertainty and I think that the reality is that we deal with a lot of uncertainty clinically in our diagnoses and therapies and so on. I think that one thing that is unfortunate is that there is no way to really discuss that uncertainty with other people and I think that often we're not honest about the uncertainty that we are experiencing because it's not perceived to be—you're not a very good doctor if you actually are not able to manage that or appear confident despite the uncertainty.* [participant 1]

Second Round Interviews With a Wireframe

Based on the thematic analysis, a wireframe was developed (see Multimedia Appendix 1), which is described as an electronic medical record plugin that could exist within the current physician workflow. The wireframe contains three main components aimed to support user needs. The "Past Patients" page allows physicians to see an automated list of patients they have previously cared for alongside data points that focus on postcare assessment and team comparisons. The "Collaboration" page allows users to speak to other physicians who may have cared for the patient, providing the opportunity for the user to ask follow-up questions or explore other inquiries. The "Learning" page allows users to find information in one consolidated place, with an additional option to take notes so information can be easily revisited.

Past Patients

Participants endorsed the inclusion of information about readmissions, additionally asking that time to readmission (eg, less than or greater than 30 days) and the readmission diagnosis be included. Participants appreciated the idea that this could be created automatically and reiterated frustrations that they are currently doing much of this work manually. Participants felt strongly about the comparative statistics as being interesting and something they would want to see, although they acknowledged there would be caveats that could influence the accuracy of these numbers.

Most participants felt positive about diagnosis but brought up a concern about trusting the diagnosis chosen, as this item can be inaccurate or a patient can have many diagnoses. Participants thought that the ability to filter through past patients would be helpful for educational purposes when looking for an example of a specific kind of patient, but that they would not likely use it in their clinical workflow.

Participants found the most recent physician and length of stay data to be the least helpful. They showed preference for finding the person themselves in the chart and only having interest in that person if it was someone they knew. For length of stay, they felt that the data would be more prone to uncontrollable variability, making the measure difficult to use.

Most participants agreed that the name, age, diagnosis, and discharge date would be their top data points for finding a past patient. Participants thought that mortality data could be added to the page.

Collaboration

Participants generally expressed that they would be unlikely to actively use the page, citing already having too many existing methods with which to communicate with other providers, a dislike of the idea of going into the medical record during off-service time to message others, and again reiterating a discomfort from earlier interviews about reaching out to others that they are less familiar with. They also worried about the potential for this chat to be monitored or used against them in some way later on.

Some participants' comments spoke to potential benefits such as having a safe, encrypted method of messaging, and asking if they would be able to attach a patient chart or specific note to a message. This would allow them to more accurately ask questions while maintaining appropriate security.

Learning

Participants expressed appreciation of having information combined into a single interface, and especially liked that they could have the information side by side with their note-writing interface, allowing them to more directly cite and annotate work notes. Although not always specifically directed toward the note-taking feature, they all expressed interest in being able to somehow note, flag, or "save for later" articles they come across. Participants expressed mixed feelings about the potential for an algorithm to automatically update feeds to match current patient issues, citing skepticism that it would be able to show them what they were really looking for. Instead, they endorsed having this page as a place to check when needing new information or passively flip through in less busy moments. Participants expressed concern that because of integration into the electronic medical record, this would not be easily accessible when not at work. The concern was rooted in the idea that coming across an article or idea can happen at any moment and would ideally be put into a readily accessible repository, which may not be the case for the wireframe.

Discussion

Principal Findings

In relation to previous self-assessment models developed in the clinical setting [7,30], our findings are consistent in illustrating that hospitalist clinicians rarely receive formal, regular, and structured feedback about their performance. However, in contrast to these prior models, our findings show that hospitalist physicians spend greater amounts of time specifically using the electronic medical record to perform self-assessment and are less reliant on peers, patients, and structural educational programs. Because the electronic medical record contains the most data available in a single place, is readily accessible to clinicians, and is a familiar and trusted platform, it is an optimal medium to explore for providing feedback to clinicians and supporting their efforts of self-assessment.

The electronic medical record is a resource easily available to hospitalists almost anywhere, providing a large array of information to use to reflect on their work. Even so, we found that the current implementation of the electronic medical record does not quickly facilitate the self-assessment and learning desired by clinicians. The electronic medical record is built for real-time care as opposed to retrospective use. It does not present summative or aggregate data, making it hard to facilitate longitudinal learning or draw conclusions based on multiple cases. We found that clinicians use time-intensive workarounds to assess themselves but often give up on these. In the evaluation of our proposed app, we found that reviewing past patients can be triggered by certain objective markers such as readmission or mortality, or subjective markers such as personal connections to certain patients or curiosity about clinically challenging cases. Given that established models and tools for peer-to-peer feedback, audit and feedback, and patient feedback exist, there is an opportunity for an app to facilitate electronic medical record–mediated self-assessment to improve the quality and standardization of this assessment [13,37,38]. Interviews of experience with our wireframe illustrate that there are some simple features that could quickly provide a strong foundation for clinician self-assessment, such as supporting a baseline database of past patients for a physician to review and organize. Although other main pages in the wireframe, such as Collaboration and Learning, led to interesting discussions, discerning clear value from these will need further exploration.

In considering whether or not to develop this kind of app, institutions should consider how clinicians are already actively looking for this information and creating their own inefficient workarounds. Developing a growth mindset culture or other positive culture around feedback is a key foundation to any intervention to combat the interpretation of feedback more negatively than it is intended [16]. Clinicians should do their best to not feel isolated by their challenges in self-assessment. They can see their workarounds and added efforts as an opportunity to work with technologists and institutions to build solutions. Technologists are challenged to find improved ways to represent this information and combat the generally negative perceptions, highlighting information that can lead to learning points. Positively, the information that clinicians have the most interest in should be easily accessible in the existing data collected. Multimedia Appendix 2 provides more detailed recommendations for institutions, clinicians, and technologists in developing this app in relation to our themes and subthemes.

We believe this is an optimal time for such an intervention. Some health care systems have embraced the learning health system model and have found ways to use the electronic medical record as a key tool in this practice, but challenges remain as the number of possible interventions is immense [39-41]. We believe that this app could engage more users to be aware of their performance and help them make more effective adjustments on their own, as opposed to relying on system-wide changes. With changes in interoperability standards and the proliferation of apps developed to integrate with electronic medical records, creation of such a tool is more technically reasonable than in the past [42,43]. We believe that following

an iterative co-design approach or participatory design approach will help lead to the best outcomes for this app [44,45].

Future Directions

Initial user research has highlighted the electronic medical record as a promising starting point for a feedback platform, and we hope to build on our current work to explore this potential. Although the medical record is central to current users, other approaches with mobile devices, apps, or other technologies could add valuable dimensions not yet captured in our current work. These other technologies will face the challenges of security, interoperability, and data accessibility to a greater degree than the electronic medical record. Applying and adapting frameworks based on behaviors such as metacognitive awareness that were demonstrated by clinicians in these interviews could also help create a more useful platform [46,47]. We hope that starting with a solid foundation of well-liked features can quickly become a branching point for new ideas and areas of focus.

Limitations

This work comes with certain limitations. Although almost all participants had worked at more than one institution in their careers, all of the participants were currently working in the same large, urban hospital system. This may lead to additional nuances that would be relevant in a more generalized population that were not revealed in this population. Participants skewed toward being less experienced as attending physicians, which may result in a bias in perspectives on feedback. Although newer clinicians may be more actively looking for feedback, more experienced clinicians may have different perspectives that were not fully captured in this work. Future work could further explore cultural differences across institutions and clinicians.

Conclusion

Our work identifies gaps and challenges in the current feedback and learning systems of hospitalist physicians. We used a qualitative approach to interview and extract themes relevant to the feedback and self-assessment of hospitalist physicians. Based on such information, we outline a gap in current apps and provide recommendations for institutions, clinicians, and technologists on how they could approach building an app to facilitate self-assessment and feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wireframe developed for second-round interviews and descriptions of each page within the wireframe.
[DOCX File , 951 KB - jmir_v22i11e23299_app1.docx]

Multimedia Appendix 2
Relevant takeaway items for each theme and subtheme, directed toward institutions, clinicians, and technologists.
[DOCX File , 17 KB - jmir_v22i11e23299_app2.docx]

References

1. ABIM Foundation. American Board of Internal Medicine, ACP-ASIM Foundation. American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. Ann Intern Med 2002 Feb 05;136(3):243-246. [doi: 10.7326/0003-4819-136-3-200202050-00012] [Medline: 11827500]

2. Li ST, Paterniti DA, Co JPT, West DC. Successful self-directed lifelong learning in medicine: a conceptual model derived from qualitative analysis of a national survey of pediatric residents. Acad Med 2010 Jul;85(7):1229-1236. [doi: 10.1097/ACM.0b013e3181e1931c] [Medline: 20592521]

3. Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. Acad Med 2013 Aug;88(8):1088-1094. [doi: 10.1097/ACM.0b013e31829a3b2b] [Medline: 23807109]

4. ACGME Common Program Requirements (Residency). Accreditation Council for Graduate Medical Education. 2018. URL: https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2019.pdf [accessed 2020-05-29]

5. Davis N, Davis D, Bloch R. Continuing medical education: AMEE Education Guide No 35. Med Teach 2008;30(7):652-666. [doi: 10.1080/01421590802108323] [Medline: 18777424]

6. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 2006 Sep 06;296(9):1094-1102. [doi: 10.1001/jama.296.9.1094] [Medline: 16954489]

7. Sargeant J, Armson H, Chesluk B, Dornan T, Eva K, Holmboe E, et al. The processes and dimensions of informed self-assessment: a conceptual model. Acad Med 2010 Jul;85(7):1212-1220. [doi: 10.1097/ACM.0b013e3181d85a4e] [Medline: 20375832]

8. LaDonna KA, Ginsburg S, Watling C. "Rising to the Level of Your Incompetence": What Physicians' Self-Assessment of Their Performance Reveals About the Imposter Syndrome in Medicine. Acad Med 2018 May;93(5):763-768. [doi: 10.1097/ACM.0000000000002046] [Medline: 29116983]

9. Gimeno García AZ. Factors influencing colorectal cancer screening participation. Gastroenterol Res Pract 2012;2012:483417. [doi: 10.1155/2012/483417] [Medline: 22190913]

10. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: a systematic review. Clin Chem Lab Med 2015 Feb;53(2):157-183 [FREE Full text] [doi: 10.1515/cclm-2014-0778] [Medline: 25263310]

11. Elligsen M, Walker SAN, Pinto R, Simor A, Mubareka S, Rachlis A, et al. Audit and feedback to reduce broad-spectrum antibiotic use among intensive care unit patients: a controlled interrupted time series analysis. Infect Control Hosp Epidemiol 2012 Apr;33(4):354-361. [doi: 10.1086/664757] [Medline: 22418630]

12. Dinescu A, Fernandez H, Ross JS, Karani R. Audit and feedback: an intervention to improve discharge summary completion. J Hosp Med 2011 Jan;6(1):28-32 [FREE Full text] [doi: 10.1002/jhm.831] [Medline: 21241038]

13. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jun 13(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]

14. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci 2014 Jan 17;9:14 [FREE Full text] [doi: 10.1186/1748-5908-9-14] [Medline: 24438584]

15. Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. BMC Health Serv Res 2016 Jul 13;16:260 [FREE Full text] [doi: 10.1186/s12913-016-1486-3] [Medline: 27412170]

16. Ramani S, Könings KD, Ginsburg S, van der Vleuten CPM. Twelve tips to promote a feedback culture with a growth mind-set: Swinging the feedback pendulum from recipes to relationships. Med Teach 2019 Jun;41(6):625-631. [doi: 10.1080/0142159X.2018.1432850] [Medline: 29411668]

17. Dannefer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, et al. Peer assessment of professional competence. Med Educ 2005 Jul;39(7):713-722. [doi: 10.1111/j.1365-2929.2005.02193.x] [Medline: 15960792]

18. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med Care 2001 Aug;39(8 Suppl 2):II2-II45. [Medline: 11583120]

19. Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. Med Educ 2012 Feb;46(2):201-215. [doi: 10.1111/j.1365-2923.2011.04137.x] [Medline: 22239334]

20. Parakh K. Assessment in medical education. N Engl J Med 2007 May 17;356(20):2108; author reply 2110. [doi: 10.1056/NEJMc070466] [Medline: 17507715]

21. Regan-Smith M, Hirschmann K, Iobst W. Direct observation of faculty with feedback: an effective means of improving patient-centered and learner-centered teaching skills. Teach Learn Med 2007;19(3):278-286. [doi: 10.1080/10401330701366739] [Medline: 17594224]

22. Bowen JL, Ilgen JS, Regehr G, Cate OT, Irby DM, O'Brien BC. Reflections From the Rearview Mirror: Internal Medicine Physicians' Reactions to Clinical Feedback After Transitions of Responsibility. Acad Med 2019 Dec;94(12):1953-1960. [doi: 10.1097/ACM.0000000000002831] [Medline: 31192795]

23. Baines R, Regan de Bere S, Stevens S, Read J, Marshall M, Lalani M, et al. The impact of patient feedback on the medical performance of qualified doctors: a systematic review. BMC Med Educ 2018 Jul 31;18(1):173 [FREE Full text] [doi: 10.1186/s12909-018-1277-0] [Medline: 30064413]

24. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. BMC Med Educ 2014 Apr 11;14:76 [FREE Full text] [doi: 10.1186/1472-6920-14-76] [Medline: 24725268]

25. Geissler KH, Friedberg MW, SteelFisher GK, Schneider EC. Motivators and barriers to using patient experience reports for performance improvement. Med Care Res Rev 2013 Dec;70(6):621-635. [doi: 10.1177/1077558713496319] [Medline: 23877957]

26. Farrington C, Burt J, Boiko O, Campbell J, Roland M. Doctors' engagements with patient experience surveys in primary and secondary care: a qualitative study. Health Expect 2017 Jun;20(3):385-394 [FREE Full text] [doi: 10.1111/hex.12465] [Medline: 27124310]

27.   Zgierska A, Rabago D, Miller MM. Impact of patient satisfaction ratings on physicians and clinical care. Patient Prefer Adherence 2014;8:437-446. [doi: 10.2147/PPA.S59077] [Medline: 24729691]

28.   Friedberg MW, SteelFisher GK, Karp M, Schneider EC. Physician groups' use of data from patient experience surveys. J Gen Intern Med 2011 May;26(5):498-504 [FREE Full text] [doi: 10.1007/s11606-010-1597-1] [Medline: 21161419]

29.   Eden AR, Hansen E, Hagen MD, Peterson LE. Physician Perceptions of Performance Feedback in a Quality Improvement Activity. Am J Med Qual 2018;33(3):283-290. [doi: 10.1177/1062860617738327] [Medline: 29088919]

30.   Lockyer J, Armson H, Chesluk B, Dornan T, Holmboe E, Loney E, et al. Feedback data sources that inform physician self-assessment. Med Teach 2011;33(2):e113-e120. [doi: 10.3109/0142159X.2011.542519] [Medline: 21275533]

31.   Holroyd-Leduc JM, Lorenzetti D, Straus SE, Sykes L, Quan H. The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence. J Am Med Inform Assoc 2011;18(6):732-737 [FREE Full text] [doi: 10.1136/amiajnl-2010-000019] [Medline: 21659445]

32.   Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]

33.   Dicicco-Bloom B, Crabtree BF. The qualitative research interview. Med Educ 2006 Apr;40(4):314-321. [doi: 10.1111/j.1365-2929.2006.02418.x] [Medline: 16573666]

34.   Yin RK. Qualitative research from start to finish, 2nd edition. New York: Guilford Publications; Jul 08, 2015.

35.   Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]

36.   Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: The Sage handbook of qualitative research in psychology. Los Angeles: Sage; Jun 30, 2017:17-37.

37.   Ossenberg C, Henderson A, Mitchell M. What attributes guide best practice for effective feedback? A scoping review. Adv Health Sci Educ Theory Pract 2019 May;24(2):383-401. [doi: 10.1007/s10459-018-9854-x] [Medline: 30284067]

38.   Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. Med Teach 2012;34(10):787-791. [doi: 10.3109/0142159X.2012.684916] [Medline: 22730899]

39.   Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med 2012 Aug 07;157(3):207-210. [doi: 10.7326/0003-4819-157-3-201208070-00012] [Medline: 22868839]

40.   Horwitz LI, Kuznetsova M, Jones SA. Creating a Learning Health System through Rapid-Cycle, Randomized Testing. N Engl J Med 2019 Sep 19;381(12):1175-1179. [doi: 10.1056/NEJMsb1900856] [Medline: 31532967]

41.   Krumholz HM, Terry SF, Waldstreicher J. Data Acquisition, Curation, and Use for a Continuously Learning Health System. JAMA 2016 Oct 25;316(16):1669-1670. [doi: 10.1001/jama.2016.12537] [Medline: 27668668]

42.   Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]

43.   Azar II A. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Federal Register: The Daily Journal of the United States Government. 2020. URL: https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification [accessed 2020-05-29]

44.   Yoo DW, Birnbaum ML, Van Meter AR, Ali AF, Arenare E, Abowd GD, et al. Designing a Clinician-Facing Tool for Using Insights From Patients' Social Media Activity: Iterative Co-Design Approach. JMIR Ment Health 2020 Aug 12;7(8):e16969 [FREE Full text] [doi: 10.2196/16969] [Medline: 32784180]

45.   Spinuzzi C. The methodology of participatory design. Tech Commun 2005;52(2):163-174.

46.   Schraw G. Promoting General Metacognitive Awareness. In: Hartman HJ, editor. Metacognition in Learning and Instruction. Neuropsychology and Cognition. Dordrecht: Springer; 2001:3-16.

47.   Schraw G. A conceptual analysis of five measures of metacognitive monitoring. Metacogn Learn 2008 Oct 24;4(1):33-45. [doi: 10.1007/s11409-008-9031-3]

XSL·FO
RenderX

<u>Original Paper</u>

Interactive Clinical Avatar Use in Pharmacist Preregistration Training: Design and Review

Jessica Thompson[1], MPharm, PhD; Simon White[1], BPharm, PhD; Stephen Chapman[1], BPharm, PhD

School of Pharmacy and Bioengineering, Keele University, Keele, United Kingdom

Corresponding Author:
Jessica Thompson, MPharm, PhD
School of Pharmacy and Bioengineering
Keele University
Hornbeam Building
Keele,
United Kingdom
Phone: 44 1782 733983
Email: j.f.thompson@keele.ac.uk

Abstract

Background: Virtual patients are interactive computer-based simulations that are being increasingly used in modern health care education. They have been identified as tools that can provide experiential learning and assessment in a standardized and safe environment. However, the study of high-fidelity virtual patients such as interactive clinical avatars within pharmacy is limited.

Objective: The aim of this paper is to describe the design and review of three interactive clinical avatar simulations as part of pharmacist preregistration training.

Methods: A multistep design approach was taken to create interactive clinical avatar simulations on the topics of emergency hormonal contraception (EHC), calculation of renal function, and childhood illnesses. All case studies were reviewed by registered pharmacists to establish content and face validity. The EHC case study and data collection questionnaire were also reviewed by a purposive sample of preregistration trainees and newly qualified pharmacists. The questionnaire used Likert ranking statements and open-ended questions to obtain users' feedback on the design, usability, and usefulness of the interactive clinical avatars as learning tools. Descriptive statistics and content analysis were undertaken on the data.

Results: Ten preregistration trainees and newly qualified pharmacists reviewed the EHC interactive clinical avatars and data collection questionnaire. The data collection questionnaire was associated with a Cronbach alpha=.95, demonstrating good reliability. All three interactive clinical avatar simulations were reported as usable and appropriately designed for preregistration training. Users perceived they were developing skills and knowledge from the simulations. The high-fidelity nature of the avatars and relevance of the simulations to real-life practice were reported as aspects that encouraged the application of theory to practice. Improvements were suggested to ensure the simulations were more user-friendly.

Conclusions: The design and creation of the interactive clinical avatar simulations was successful. The multistep review process ensured validity and reliability of the simulations and data collection questionnaire. The in-depth explanation of the design process and provision of a questionnaire may help widen the use and evaluation of interactive clinical avatars or other simulation tools in pharmacy education. The interactive clinical avatars were reported as novel learning tools that promoted experiential learning and allowed users to feel like they were engaging in real-life scenarios, thus developing transferable knowledge and skills. This may be potentially beneficial for many health care training courses as a way to provide standardized experiences promoting active learning and reflection.

Introduction

Virtual patients are defined as "a specific type of computer-based program that simulates real-life clinical scenarios [in which] learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions" [1]. Virtual patient is an umbrella term that includes computer-based tools such as still photos, video clips, avatars, and immersive virtual reality simulations. The key element for a tool to be described as a virtual patient is that the simulation outcome must be dependent on user input, their prime differentiation from other e-learning tools [2].

An interactive clinical avatar is classified as a high-fidelity form of virtual patient [3]. The avatar is a computer-generated, 3-dimensional (3D) animation that represents a patient or health care professional. Avatars can be either asynchronous, accessed anytime and providing responses without the requirement for a tutor online, or synchronous, active in real time and requiring input from multiple users or educators [4,5].

Interactive clinical avatars have been evaluated less frequently in pharmacy education and training than in other health care professions and less frequently than other forms of virtual patients [2,6,7]. This may be as a result of potential barriers associated with their production and use including information technology skills, cost, and time constraints [6,8,9]. Evaluations of asynchronous avatars in pharmacy education have demonstrated significant positive outcomes in the development of students' knowledge, skills, and confidence to interact with real patients [10-15]. The majority of this research has used self-reporting scales based on individuals' perceptions to evaluate outcomes. The drawbacks of such scales is their inherent subjectivity; only individuals' perceptions can be reported, with the appropriateness of statistical analysis to determine significance greatly debated in the literature [16,17].

Pharmacy preregistration training in the United Kingdom is a workplace-based training year predominantly occurring in a community or hospital pharmacy. Performance standards must be met during the preregistration training year before trainees can take the preregistration examination and be registered as a pharmacist. Experiences of preregistration trainees can vary considerably and are dependent on multiple factors including sector of training (hospital vs community), training site, tutor support, and individual initiative [18,19]. Notable differences in the quality of preregistration training have resulted in a disparity in preregistration examination pass rates between the sectors of training [20]. The role of a pharmacist in the United Kingdom is evolving, and training needs to adapt to ensure all individuals qualify with appropriate knowledge and skills to provide safe and effective patient care [21].

Simulation has the potential to reduce variation in preregistration training by providing standardized experiences, thus ensuring a more controlled development of trainee competence. Miller's triangle is the most commonly used model to measure competency development in pharmacist preregistration training [22]. A benefit of simulation, especially high-fidelity simulation, is the promotion of experiential learning, which aids user progression to the "shows how" level of competence, initially proposed as difficult to reach without real-world experience [23,24]. Immersive simulations provide a safe environment for individuals to practice specific skills or knowledge without risk of harm to a real patient [25]. Evaluation of high-fidelity simulations is essential moving forward with the increasing numbers of pharmacy students entering university, changing role of a pharmacist, and problems with standardizing placements [26]. Experience alone may not be enough for individuals to obtain mastery of clinical skills [27], and simulation-based learning can be instrumental in promoting experiential, situated learning and bridging the gap between theory and practice without differences in educational outcomes [28].

Only one study has evaluated interactive clinical avatars in pharmacist preregistration training, leading to the creation of 3 interactive clinical avatar case studies covering different competencies [10]. The interactive clinical avatar software used has 3 key parts associated with and essential to its design: an electronic database containing the avatar's responses, a computer-generated graphic, and a system to link the two together [29]. The brain of the avatar is the electronic database based on a modified Markov model design [30], which uses a decision tree to map the progress of a case. As a user progresses through a case, the database monitors the path through the decision tree and collates positive and negative feedback to be given at the end of the simulation [31]. The body of the avatar is a computer-generated graphic of a 3D character. The final part of the system is the heart, which carries information from the brain to the body, allowing a real-time, immediate response.

The aim of this paper is to describe the design and review of three interactive clinical avatar simulations that are part of pharmacist preregistration training.

Methods

Intervention Development

A multistep design approach was taken.

Case Design

Case studies were designed and created by the research and digital development teams at Keele University School of Pharmacy and Bioengineering from a review of the literature, the preregistration syllabus, and discussions between the lead researcher (JT) and first year qualified pharmacists who had recently completed their training year. Interactive clinical avatar simulations were created on the topics of emergency hormonal contraception (EHC), calculation of renal function, and childhood illnesses to develop a range of knowledge and skills essential for preregistration training and future practice. Due to training variations, trainees may find it difficult to demonstrate competence in these areas. The clinical elements of the cases were based on appropriate guidelines and resources that pharmacists use in everyday practice.

The simulation on EHC supply was created because trainees completing their preregistration training in a hospital would be less likely to see or be actively involved in an EHC consultation than those in the community sector. Individuals training in a

community pharmacy are only able to observe a consultation, so a simulation was created to provide an opportunity to actively advise on the use of EHC. The calculation of renal function was incorporated into a simulation because hospital pharmacists calculate patients' renal function and adjust drug doses more often than those working in community pharmacy. The childhood illness case study was designed around measles because of its increasing prevalence in the United Kingdom and the need for pharmacists in the community and hospital sectors to have the knowledge to distinguish it from other conditions [32].

Script Design

The scripts for the interactive clinical avatar cases were designed to meet predefined learning objectives and ensure equivalent user experiences. The research team worked in consultation with a community pharmacist on the EHC case and a hospital pharmacist in the renal function case. The scripts were reviewed and amended according to the principles of constructive alignment in face-to-face meetings and through email until realistic and accurate simulations were created that aligned with the intended learning outcomes [33]. A decision tree of the script for each case was developed that established key points and decisions. Each case had an ideal pathway, and decisions were categorized as more or less favorable compared with this (Figure 1).

Figure 1. Markov model decision tree for the emergency hormonal contraception simulation.
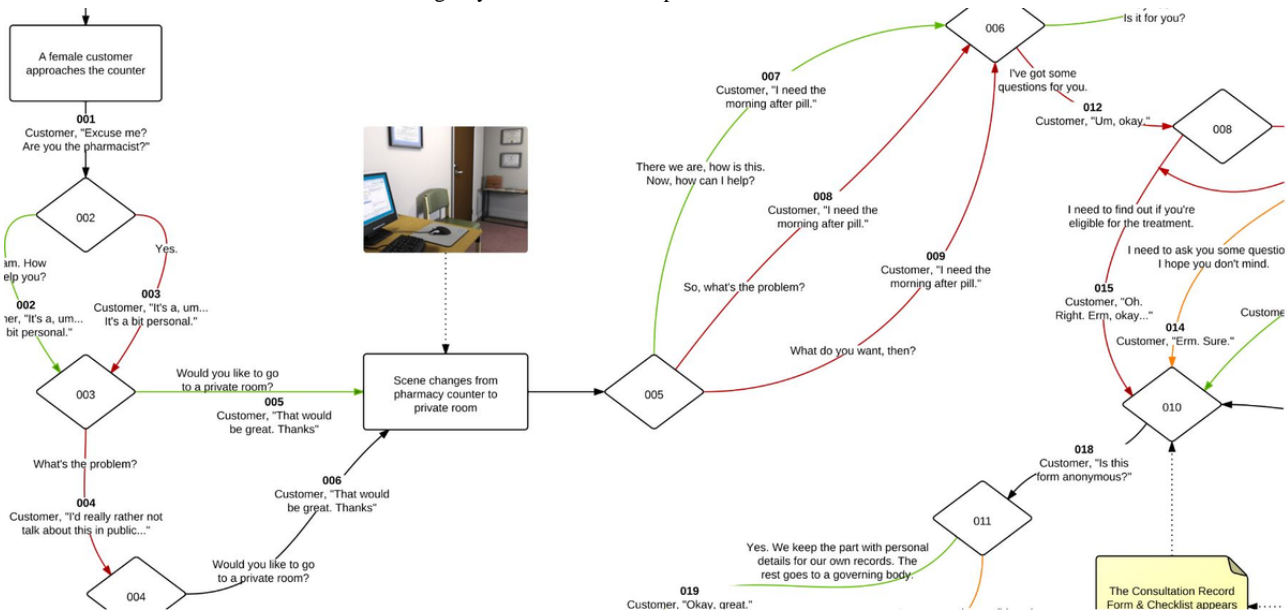
Different script styles were created depending on the intended learning outcomes of the simulation. Standard navigation through virtual patient simulations is via predefined menu inputs [34]; however, different inputs were used in the design to widen the scope of the interactive clinical avatars used and evaluate user preferences in a larger study [35]. Multiple-choice inputs were selected for communication skill development, whereas free-text inputs were used to encourage knowledge development [36-38]. The interactive clinical avatars were designed to provide feedback to users. This feedback, spoken by the avatar and provided in textual format, specifically laid out what a user had done well and how they could improve. This provision was selected to encourage a reflective approach to learning, as was the ability for a user to repeat the simulation and change their path to receive a different patient outcome and feedback.

Avatar Design

The scripts and avatars were designed to be realistic but not to the detriment of falling into the uncanny valley as described by Mori [39], who suggested that objects portraying human characteristics may be viewed positively up to a certain point beyond which the degree of visual similarity to real humans becomes unsettling and can trigger negative thoughts [40-42]. All avatars were designed to express humanistic characteristics

through simple body language, movements, and prerecorded voice replies. As an example, the avatar in the EHC simulation was animated to blush when asking for the morning after pill, as may happen in real practice due to the sensitive nature of the topic.

The simulations were stored online and created using standard HTML. Javascript functions were used to handle page logic, including processing button clicks for both the free-text input and multiple-choice questions. For the interactive clinical avatar simulations using free-text inputs, once the Speak to Patient button was clicked, the text was sent to a web service hosted on the same server to process the text. This returned a code to the site processed on the client side via JavaScript (using the jQuery library) to determine what animation to play and how the case should continue. When multiple-choice inputs were used, processing occurred client side; skipping the web service step. The avatars were modeled, textured, and animated using Maya 3D software (Autodesk Inc). Still images were rendered using mental ray (Autodesk Inc) and then composited using After Effects (Adobe) to create the final MP4 animation files. The avatars' verbal responses were audio recordings of a voice actor who was chosen to match the avatars' characteristics.

The avatar in the EHC simulation was a young woman who presented to a community pharmacy requesting the morning after pill (Figure 2). The user had the option to invite the avatar into a consultation room to ensure a private, confidential conversation, and the simulation background would change to show this. The simulation used a mix of multiple-choice and free-text inputs. To aid the consultation and mirror real life, a consultation form was built into the simulation and included as a separate tab, which could be opened and completed by the user as they progressed through the consultation to help them make a decision on the appropriateness of EHC supply.

Figure 2. Screenshot of the emergency hormonal contraception avatar simulation.
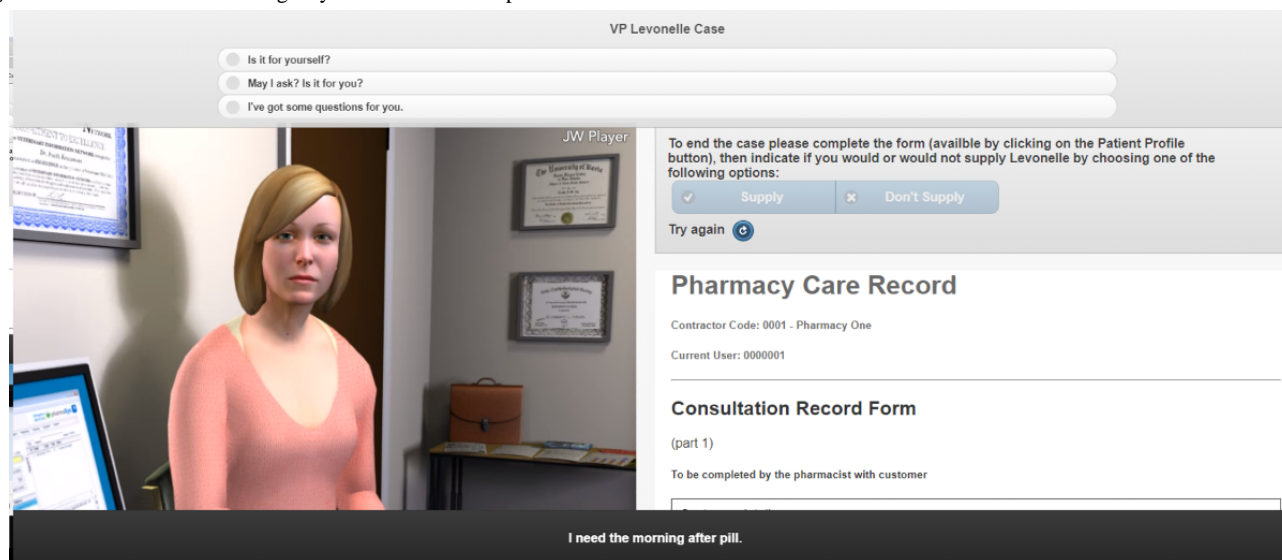
In the renal function simulation, users played the role of a hospital pharmacist and interacted with an avatar of a doctor on a hospital ward via multiple-choice input (Figure 3). The simulation integrated hospital notes and a drug chart that could be accessed throughout the simulation and contained the same sort of information as would be expected in a real set of hospital notes, including test results and medication history. Users were required to calculate the patients' renal function using the information provided and adjust drug doses accordingly to help improve their calculation and clinical skills.

Figure 3. Screenshot of the renal function avatar simulation.
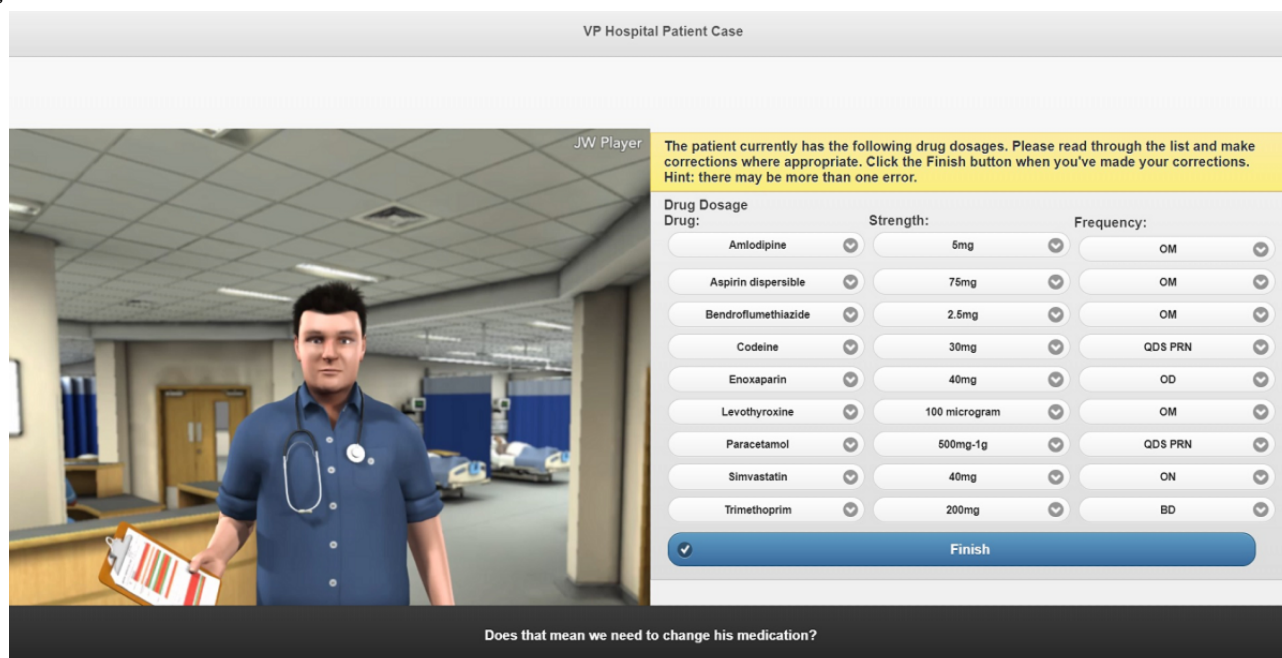
The childhood illness simulation used two avatars: a grandmother and her grandson (Figure 4). The simulation was set in a community pharmacy, and users interacted with both avatars via free-text input. The grandson avatar presented with measles. When users asked to look at the rash, an enlarged image of the child's arm appeared showing a more detailed image of the rash associated with measles; similarly, if the user requested to look inside the child's mouth, an enlarged image of the oral cavity appeared showing Koplik spots. These images remained available throughout the simulation and demonstrated the

high-level fidelity of the technology. The user was directed to make a diagnosis, and upon success, they were directed to recommend appropriate licensed medicines and self-care advice.

Avatars were not based on real patients, and any resemblance to a person, living or dead, is coincidental.

A YouTube video has been created as a demonstration of the EHC avatar simulation [43] (Multimedia Appendix 1).

Figure 4. Screenshot of the childhood illness avatar simulation.
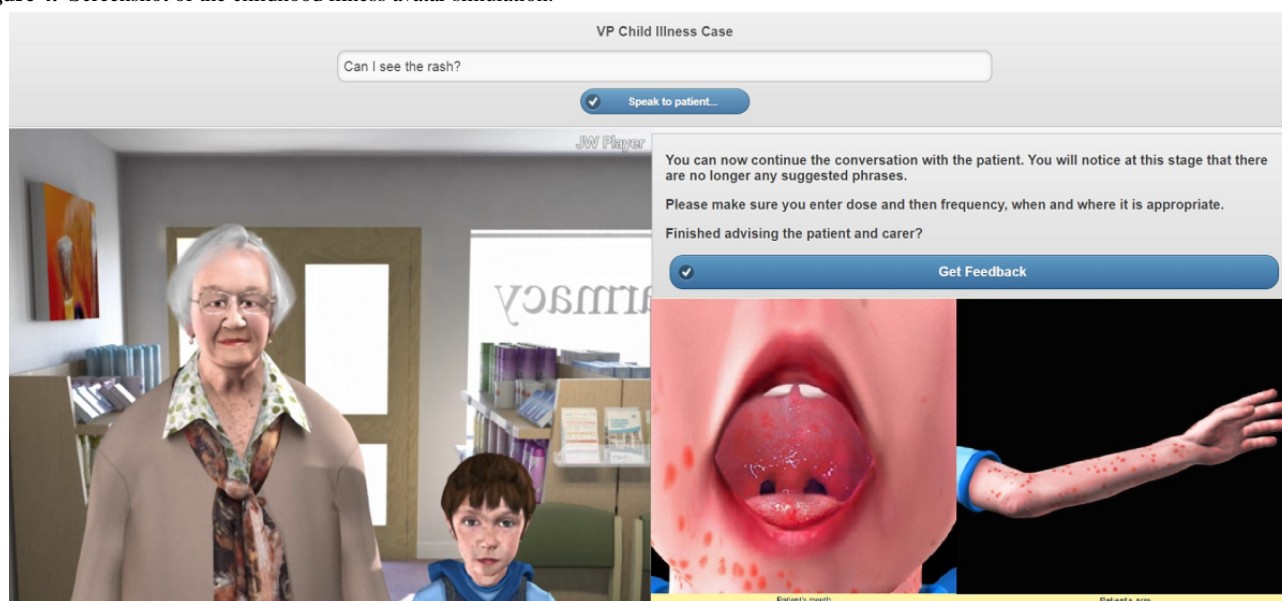
Interactive Clinical Avatar Evaluation

Design

During the initial design phase, an internal review took place within the research team with advice from the digital development team and feedback from registered pharmacists. The interactive clinical avatar simulations were tested by pharmacists within the School of Pharmacy and Bioengineering at Keele University; all members of staff who were pharmacists were contacted, including those who had assisted with the initial script design. Links to the interactive clinical avatar simulations were sent via email, and pharmacists were asked to provide electronic or face-to-face feedback on any areas of the case studies, including clinical and pedagogical aspects.

In addition, the EHC case study and data collection questionnaire were reviewed by a sample of preregistration trainees and newly qualified pharmacists. Ethical approval was granted by the Keele University Ethical Review Panel.

Sampling and Recruitment

A purposive sample of 30 preregistration trainees and newly qualified pharmacists who were Keele University alumni (n=10) or members of the British Pharmaceutical Students Association (n=20) were invited to participate over social media. Consideration was given to the sample size based on the pool of available participants and the required sample size for the wider study [44]. Previous literature has identified a range of sample sizes for studies designed to review data collection materials, with a minimum of 10 established as appropriate to evaluate adequacy of instrumentation [45]. A range of views was sought for this evaluation, and thus recruitment included individuals who had studied at different universities, were in different stages of their education, and were currently

completing or had completed their preregistration training in a community or a hospital pharmacy. The study used a sample that best represented those with knowledge of the research topic, in this case current preregistration trainees and individuals who had just finished preregistration training. Only one round of recruitment occurred, as saturation of themes was reached [46].

Data Collection

Participants were asked to complete the interactive clinical avatar case study on the knowledge and law of supplying EHC and a questionnaire to gather their feedback on the simulation and the data collection tool. Evaluative questionnaires are an established method to compare thoughts on different learning tools [47,48], but a review of the literature identified a lack of validated instruments to evaluate interactive clinical avatars as learning tools [49]. A data collection questionnaire was therefore created for this research by adapting questions from validated evaluation instruments found in the educational literature to provide a tool to evaluate perceptions of virtual patients in pharmacy education for future studies [50-52].

This questionnaire consisted of 20 Likert scale ranking statements (1= strongly disagree, 2= disagree, 3= undecided, 4= agree, 5= strongly agree) and a series of open-ended questions and was subject to measurements of reliability and validity. Review by the supervisory team, practicing pharmacists, preregistration trainees, and newly qualified pharmacists ensured its content, face validity, and internal consistency [53-55].

Data Analysis

Quantitative data from the Likert scale were analyzed descriptively. Qualitative data from the open-ended questions were subject to content analysis [56]. All comments provided by participants were imported into an Excel spreadsheet

(Microsoft Corp) for coding. The content analysis process involved the identification, organization, and indexing of comments to provide frequencies of coded themes [57]. The lead researcher (JT) conducted the initial content analysis. These codes were reviewed by the research team and discussed to reach consensus, providing researcher triangulation. Cronbach coefficient alpha measured internal reliability of the data collection tool [58].

Results

Demographic Results

Of 30 invited, 10 participants responded and reviewed the EHC interactive clinical avatar and data collection questionnaire: 4 were Keele University alumni and 6 studied at other universities and were members of the British Pharmaceutical Students Association. Four of the participants were preregistration trainees, and 6 were first-year qualified pharmacists; 7 of the participants were female and 3 were male. All participants were aged 20 to 25 years. Participants were of White British (7/10), Indian (1/10) and Chinese (2/10) ethnicities. An equal number of participants were from hospital and community practice. A greater proportion of participants (6/10) reported no previous experience with interactive clinical avatar simulations.

Questionnaire Validation

Initial reliability testing of the instrument consisted of establishing content and face validity. The research team as pharmacists had first-hand knowledge of the construct of interest from practice, and they could advise on the questionnaire content and language. The review of the questionnaire by three qualified pharmacists contributed to addressing the content validity of the instrument. This resulted in minor typographical changes, but the questionnaire was otherwise reported as appropriate to measure user views on interactive clinical avatar simulations.

All preregistration pharmacists and newly qualified pharmacists reported the questions and Likert statements to be clearly written and unambiguous, further confirming internal validity of the data collection tool. Cronbach α=.95, which demonstrated a high level of internal consistency suggesting that the Likert scales were reliable to be used in a bigger study [44].

Quantitative Analysis

The statements from the Likert ranking scales, their descriptions, and the associated median agreement scores are shown in Table 1. For all Likert ranking statements, participant median agreement scores were in the agree or strongly agree categories. Fewer participants ranked themselves as undecided, and only on 5 occasions did participants disagree with a statement. No participants ranked themselves as a 1 (strongly disagree) for any statement.

The fidelity of the simulation was reported as realistic by 80% (8/10) of participants, with it making them feel as if they were the pharmacist and were making the same decisions as in real practice (70% [7/10] and 90% [9/10] agreement, respectively). The simulation was reported as enjoyable (9/10, 90%), interesting (10/10, 100%), and set at the right level for preregistration training (9/10, 90%). The majority of participants reported that the simulation was easy to access (9/10, 90%) and the intended learning outcomes were easy to understand (7/10, 70%), indicating high usability as a learning tool.

The level and specificity of the feedback was reported as adequate for participants' learning needs (10/10, 100%). The development of skills for future practice from the simulation was reported by 80% (8/10) of participants, with clinical reasoning skills being developed to a greater extent than problem-solving or decision-making skills (90% [9/10] vs 80% [8/10]). There was high agreement regarding the development of and application of knowledge from using the simulation (100% [10/10] and 80% [8/10], respectively). Participants reported feeling more confident in caring for patients (9/10, 90%), collaborating with other health care professionals (8/10, 80%) and practicing as a pharmacist (8/10, 80%) after completing the simulation.

There were some differences between data received from preregistration trainees and newly qualified pharmacists. Preregistration trainees agreed or strongly agreed with the majority of Likert statements; one participant (25%) was undecided with statements 2 and 3 relating to really feeling like the pharmacist in the simulation. Two newly qualified pharmacists (33%) also ranked themselves as undecided on statement 2. Newly qualified pharmacists were reportedly more undecided as to the simulation aiding the development of skills for practice (2/6, 33%) and allowing the application of theory to practice (2/6, 33%). There were a few statements on which newly qualified pharmacists disagreed: aiding the development of problem-solving and decision-making skills (1/6, 17%), individual responsibility to learning (1/6, 17%), collaborating with patients (1/6, 17%) and health care professionals and increasing confidence for practice (2/6, 33%).

Table 1. Participant perceptions of the interactive clinical avatar simulation.

Statement number	Description	Agreement score, median (IQR)
1	Simulation provided a realistic patient simulation.	4.5 (1)
2	When completing the simulations, I felt as if I were the pharmacist caring for this patient.	5.0 (1.75)
3	When completing the simulations, I felt I had to make the same decisions as a pharmacist would in real life.	4.5 (1)
4	Simulations were interesting.	4.5 (1)
5	Simulations were enjoyable.	4.5 (1)
6	Difficulty of the simulations was appropriate for my level of training.	4.0 (0.75)
7	Feedback I received was adequate for my needs.	5.0 (1)
8	Objectives for the simulations were clear and easy to understand.	4.0 (0.75)
9	I was able to access the simulations at my convenience.	5.0 (1)
10	Simulations helped develop my clinical reasoning skills.	4.5 (1)
11	Simulations helped develop my problem-solving and decision-making skills.	4.0 (1)
12	Simulations have helped me put theory into practice.	4.5 (1)
13	I am confident I am developing the skills required in practice from the simulations.	4.0 (0.75)
14	I am confident I am gaining the knowledge required in practice from the simulations.	4.5 (1)
15	It is my responsibility to learn what I need to know from the simulations.	5.0 (1)
16	Completing the simulations has improved my confidence for the preregistration exam.	4.5 (1)
17	I feel better prepared to care for real-life patients.	4.0 (0.75)
18	I feel more confident about collaborating with patients and other health care professionals.	4.0 (0)
19	Simulations have increased my confidence about practicing as a pharmacist.	4.0 (0.75)
20	Overall, the experience has enhanced my learning.	4.5 (1)

Qualitative Analysis

Key themes emerged from content analysis of the open-ended questions on the questionnaire regarding use of the case studies as learning tools, use of the case studies in the preregistration training year, limitations of the case studies, and suggestions for improvements of the case studies. Themes and frequencies from the content analysis can be found in Multimedia Appendix 2.

Participants reported the interactive clinical avatar simulation as enjoyable and easy to use as a learning tool. It was described as providing a different format of learning, and the realism and relevance to practice of the simulation promoted experiential learning.

> *I liked that there was a selection of answers to reply to the patient, more than just a "yes" and "no"...I feel this is how I would communicate with a patient politely, and it made the experience a lot more realistic.* [P4, preregistration trainee]

One participant specifically commented that the level of feedback was helpful to identify what they did well and how they could improve, while drawing on the repeatable aspect of the simulation.

> *It helps put learnt knowledge into practice...in a safe and nonpressured environment. I could make a mistake and learn from it; learn how to do a proper,*

> *effective consultation without compromising patient care in real life.* [P2, preregistration trainee]

Participants reported that the interactive clinical avatar could be a useful learning tool in preregistration training; it was primarily thought of as an individual revision aid or a group learning tool. A small number of participants commented on the use of interactive clinical avatar as a tool to help prepare for an Objective Structured Clinical Examination (OSCE).

> *This case study would fit well in an OSCE scenario...more realistic than a written station and you wouldn't need to bring in a [standardized patient].* [P7, newly qualified pharmacist]

Participants reported some limitations of the interactive clinical avatar. The most frequent concern related to the free-text part of the simulation.

> *At the end of the case study where I had to ask questions. I didn't realize that this section could also be used to counsel and give advice to the patient. I felt this wasn't made clear.* [P3, newly qualified pharmacist]

The free-text entry was also associated with software recognition problems.

> *It was difficult to enter information as the programme often didn't recognize these. It became frustrating*

thinking how to reword things to make the programme recognize them. [P5, newly qualified pharmacist]

Four improvements were suggested: providing an example simulation, including a help button or email address for technical issues, including key learning points at the end of the case study, and increasing the size of the question bank (most commonly suggested).

Have a bigger question bank so the simulation can recognize what you are asking and make it easier to progress through the case...it makes you not want to carry on with the consultation. [P9, preregistration trainee]

Discussion

Principal Findings

The aim of the study was to describe the design and review of three interactive clinical avatar simulations as part of pharmacist preregistration training. Creating the interactive clinical avatar simulations involved multiple reviews and subsequent amendments to ensure content and face validity of the cases. Further measures of concurrent or construct validity were considered and dismissed as not appropriate because of the lack of validated instruments and, therefore, the inability for findings to be correlated with those from another instrument. A sample of preregistration trainees and newly qualified pharmacists reviewed the EHC simulation and reported it to be an effective learning tool, suggesting improvements to increase its usability. The wording, relevance, and difficulty of the interactive clinical avatar were reported as being appropriate with no changes required. The data collection questionnaire was found to be reliable, as indicated by the high Cronbach alpha score.

There has been only one previous study evaluating interactive clinical avatars in preregistration pharmacy training [10], which focused on improving users' knowledge and confidence of a single community pharmacy service. In contrast to our study, there were few specifics on the simulation design process or validity of the self-rating scale. Previous literature reviews [2,6,7,59,60] concluded the need for further research to be undertaken on the role of virtual patients in postgraduate and continuing pharmacy education. The design of the interactive clinical avatars in this paper goes some way to filling this gap in current research and provides a rationale for a combination of input styles to determine the range of skills and knowledge that can be developed from these simulations [35,44].

This study indicated a high level of satisfaction by preregistration trainees and newly qualified pharmacists with the use of interactive clinical avatars as a learning tool, mirroring the wider literature [12,13,61-65]. Providing engaging and enjoyable learning environments are key components of the learning process [66]. Participants in this study enjoyed completing the EHC simulation, describing it as a novel learning tool that had not been widely used in participants' undergraduate education. The increasingly digitally native generation of students justifies the creation and evaluation of these more interactive learning tools that may promote self-directed and

distance learning, potentially increasing the scope of students who can use them [67].

Participants commented on the realistic design and relevance of the scenarios to real practice, implying high levels of cognitive realism [68]. The interactivity and fidelity of the interactive clinical avatar may have added to users' immersion in the simulation and their experiential, situated learning; such learning has previously been proposed to occur only with real-life or concrete experiences [69]. Participants reported that the interactive clinical avatar provided them with a means to explore higher order skills such as clinical reasoning. In contrast, the development of more simple skills (such as problem solving or decision making) were not reported as freely, especially by newly qualified pharmacists. This is not necessarily an issue, as previous studies have found that lower order skills can be acquired through straightforward rote learning, but it remains essential to ensure that any learning tool is created with the level of learners and intended learning outcomes in mind [70]. Although virtual patients have been established as tools to develop health care professionals' clinical reasoning skills [2,8,28,71,72], little work has been conducted within the pharmacy profession or evaluating interactive clinical avatars specifically. Findings from this study also mirror the wider literature regarding knowledge development post virtual patient use, with preregistration trainees and newly qualified pharmacists reporting the development of knowledge required for future practice after completing the interactive clinical avatar simulation [10,13,73-79]. Resources that allow users to immerse themselves in a scenario and behave as they would in real practice are extremely beneficial in aiding the development of knowledge, skills, and confidence and may bridge the gap between theory and practice in preregistration training [80].

Learning tools must reflect the level of education that individuals are at and allow a spirality of content to ensure new learning is linked to previous learning and students' competence can develop in line with the difficulty of task [81]. Interactive clinical avatars may add little value when learners are at the lower levels of Bloom's taxonomy [82]; they should instead be used when knowledge is combined with skills and applied in problem-solving scenarios or when direct patient contact is not possible. This may explain why preregistration trainees were more agreeable to the simulations allowing an opportunity to put theory into practice, developing skills, improving confidence for real-life practice, and collaborating with patients and health care professionals; newly qualified pharmacists may have had this experience in real-life practice. This demonstrates that interactive clinical avatars are able to help promote progression through the levels of Miller's triangle, as per other simulation-based learning [83].

The integration of a decision tree into simulation design allows users to dynamically control the case study and assures outcomes are directly determined by user input to encourage a greater sense of realism and responsibility. Interactive clinical avatar simulations provide a safe environment for users to repeatedly practice, change the simulation outcome, receive the appropriate feedback, and reflectively learn. This was noted as being particularly beneficial in preregistration training, as it would

allow individuals to learn from experiences they may not otherwise have.

Strengths and Limitations

All interactive clinical avatar case studies were reviewed by the research team and qualified pharmacists to increase their validity. The EHC simulation was also reviewed by a sample of preregistration trainees and newly qualified pharmacists. The sample size of 10 may be considered small and findings due to chance, but the purpose was to review the design of the interactive clinical avatar simulation and data collection tool, for which data saturation was reached. All comments received were useful regarding the interactive clinical avatar design—in particular, improvements suggested related to features consistent for all three case studies (eg, providing an example interactive clinical avatar simulation, signposting for help, and listing key points relating to the topic). The improvements regarding the question bank and recognition of the avatar can also be translated across the interactive clinical avatar simulations where free-text input is used.

Perceptions of the interactive clinical avatar were obtained via a questionnaire. The self-reporting aspect did not show whether participants' knowledge or skills related to completing an EHC consultation did actually improve, only their thoughts of such.

The primary focus of the questionnaire was to obtain participant perceptions relating to the avatar simulation rather than evaluating its ability to promote knowledge or skill development; this will be part of a larger study.

Conclusions

The design and creation of the interactive clinical avatar simulations was successful. The multistep review process ensured validity and reliability of the simulations and data collection questionnaire. The in-depth explanation of the design process and provision of a questionnaire may help widen the use and evaluation of interactive clinical avatars or other simulation tools.

The interactive clinical avatars were received favorably as a novel learning tool that was reported to promote experiential learning and allow individuals to feel like they were making real-life decisions, developing transferable knowledge and skills. Improvements were suggested to increase the use of the interactive clinical avatars, and amendments will be made for further, larger scale evaluations. The use of interactive clinical avatars as learning tools can provide potential benefits to many health care training courses because they provide standardized simulations that allow active learning and reflection.

Authors' Contributions

All authors (JT, SW, SC) were involved in developing the study design and methods. JT conducted the investigation and formal analysis of the data and was responsible for writing the first draft of the manuscript. SW and SC supervised the project. All authors contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Emergency hormonal contraception interactive clinical avatar demonstration.
[MP4 File (MP4 Video), 22852 KB - jmir_v22i11e17146_app1.mp4]

Multimedia Appendix 2
Content analysis themes and frequencies.
[PDF File (Adobe PDF File), 400 KB - jmir_v22i11e17146_app2.pdf]

References

1. Effective use of educational technology in medical education—colloquium on educational technology: recommendations and guideline for medical educators. AAMC Institute for Improving Medical Education. 2007. URL: https://store.aamc.org/downloadable/download/sample/sample_id/111/ [accessed 2019-09-01]
2. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
3. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teacher 2009 Sep 09;31(8):743-748. [doi: 10.1080/01421590903124708]
4. Bell MW. Toward a definition of "virtual worlds". J Virtual Worlds Res 2008 Jul 01;1(1):1-5. [doi: 10.4101/jvwr.v1i1.283]

5. Petrakou A. Interacting through avatars: virtual worlds as a context for online education. Comput Educ 2010;54(4):1020-1027. [doi: 10.1016/j.compedu.2009.10.007]

6. Jabbur-Lopes MO, Mesquita AR, Silva LMA, De Almeida Neto A, Lyra DP. Virtual patients in pharmacy education. Am J Pharm Educ 2012 Jun 18;76(5):92 [FREE Full text] [doi: 10.5688/ajpe76592] [Medline: 22761533]

7. Richardson C, White S, Chapman S. Virtual patient technology to educate pharmacists and pharmacy students on patient communication: a systematic review. BMJ Simul Technol Enhanc Learn 2019;5(A39):1. [doi: 10.1136/bmjstel-2019-000514]

8. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]

9. Cavaco AM, Madeira F. European pharmacy students' experience with virtual patient technology. Am J Pharm Educ 2012 Aug 10;76(6):106 [FREE Full text] [doi: 10.5688/ajpe766106] [Medline: 22919082]

10. Zlotos L, Power A, Hill D, Chapman P. A scenario-based virtual patient program to support substance misuse education. Am J Pharm Educ 2016 Apr 25;80(3):48 [FREE Full text] [doi: 10.5688/ajpe80348] [Medline: 27170819]

11. Taglieri CA, Crosby SJ, Zimmerman K, Schneider T, Patel DK. Evaluation of the use of a virtual patient on student competence and confidence in performing simulated clinic visits. Am J Pharm Educ 2017 Jun;81(5):87 [FREE Full text] [doi: 10.5688/ajpe81587] [Medline: 28720915]

12. Ferrone M, Kebodeaux C, Fitzgerald J, Holle L. Implementation of a virtual dispensing simulator to support US pharmacy education. Curr Pharm Teach Learn 2017 Jul;9(4):511-520. [doi: 10.1016/j.cptl.2017.03.018] [Medline: 29233422]

13. McDowell J, Styles K, Sewell K, Trinder P, Marriott J, Maher S, et al. A simulated learning environment for teaching medicine dispensing skills. Am J Pharm Educ 2016 Feb 25;80(1):11 [FREE Full text] [doi: 10.5688/ajpe80111] [Medline: 26941437]

14. Bindoff I, Ling T, Bereznicki L, Westbury J, Chalmers L, Peterson G, et al. A computer simulation of community pharmacy practice for educational use. Am J Pharm Educ 2014 Nov 15;78(9):168 [FREE Full text] [doi: 10.5688/ajpe789168] [Medline: 26056406]

15. Berger J, Bawab N, De Mooij J, Sutter Widmer D, Szilas N, De Vriese C, et al. An open randomized controlled study comparing an online text-based scenario and a serious game by Belgian and Swiss pharmacy students. Curr Pharm Teach Learn 2018 Dec;10(3):267-276. [doi: 10.1016/j.cptl.2017.11.002] [Medline: 29764629]

16. Jamieson S. Likert scales: how to (ab)use them. Med Educ 2004 Dec;38(12):1217-1218. [doi: 10.1111/j.1365-2929.2004.02012.x] [Medline: 15566531]

17. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]

18. Guile D, Ahamed F. Modernising the pharmacy curriculum. Centre for Learning and Life Chances in Knowledge Economies and Societies. 2011. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.671.632&rep=rep1&type=pdf [accessed 2019-09-01]

19. Mills E, Blenkinsopp A, Black P. Quality management in pharmacy pre-registration training: current practice. Pharm Educ 2013;13(1):82-86 [FREE Full text]

20. The June 2015 Registration Assessment Meeting of the Council. General Pharmaceutical Council. 2015 Sep 10. URL: https://52.49.7.134/sites/default/files/2015-09-10_gphc_council_papers_public_-_post-meeting.pdf [accessed 2020-10-23]

21. John C. The changing role of the pharmacists in the 21st century. Pharm J 2018;300:1 [FREE Full text] [doi: 10.1211/PJ.2018.20204131]

22. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]

23. Smith A, Darracott R. Modernising pharmacy careers programme—review of pharmacist undergraduate education and pre-registration training and proposals for reform: report to Medical Education England Board. 2011. URL: https://tinyurl.com/y43yolkv [accessed 2019-09-01]

24. Lalley J, Miller R. The learning pyramid: does it point teachers in the right direction? Education 2007;128(1):64-79.

25. Ker J, Bradley P. Simulation in medical education. In: Swanwick T, editor. Understanding Medical Education: Evidence, Theory and Practice. 1st Edition. Hoboken: Wiley-Blackwell; 2010:164-180.

26. Badcock LJ, Raj N, Gadsby K, Deighton CM. Meeting the needs of increasing numbers of medical students: a best practise approach. Rheumatology (Oxford) 2006 Jul;45(7):799-803. [doi: 10.1093/rheumatology/kel070] [Medline: 16705052]

27. Jeffries PR, Beach M, Decker SI, Dlugasch L, Groom J, Settles J, et al. Multi-center development and testing of a simulation-based cardiovascular assessment curriculum for advanced practice nurses. Nurs Educ Perspect 2011;32(5):316-322. [doi: 10.5480/1536-5026-32.5.316] [Medline: 22029244]

28. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]

29. Bracegirdle L, Chapman S. Programmable patients: simulation of consultation skills in a virtual environment. Bio-Algorithms Med Syst 2010;6(11):111-115 [FREE Full text] [doi: 10.2478/bams]

30. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]

31. Richardson C, Chapman S, White S. Virtual patient educational programme to teach counselling to clinical pharmacists: development and proof of concept. BMJ Simul Technol Enhanc Learn 2019;5:167-169. [doi: 10.1136/bmjstel-2018-000352]

32. Public HE. National Measles Guidelines. Public Health England. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/849538/PHE_Measles_Guidelines.pdf [accessed 2020-10-23]

33. Biggs J. Enhancing teaching through constructive alignment. High Educ 1996 Oct;32(3):347-364. [doi: 10.1007/BF00138871]

34. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676. [doi: 10.2196/14676]

35. Thompson J, White S, Chapman S. Virtual patients as a tool for training pre-registration pharmacists and increasing their preparedness to practice: a qualitative study. PLoS One 2020;15(8):e0238226 [FREE Full text] [doi: 10.1371/journal.pone.0238226] [Medline: 32866197]

36. Huwendiek S, Reichert F, Bosse H, de Leng BA, van der Vleuten CPM, Haag M, et al. Design principles for virtual patients: a focus group study among students. Med Educ 2009 Jun;43(6):580-588. [doi: 10.1111/j.1365-2923.2009.03369.x] [Medline: 19493183]

37. Botezatu M, Hult H, Fors UG. Virtual patient simulation: what do students make of it? A focus group study. BMC Med Educ 2010 Dec 04;10(91):1 [FREE Full text] [doi: 10.1186/1472-6920-10-91] [Medline: 21129220]

38. Jacklin S, Maskrey N, Chapman S. Improving shared decision making between patients and clinicians: design and development of a virtual patient simulation tool. JMIR Med Educ 2018 Nov 06;4(2):e10088 [FREE Full text] [doi: 10.2196/10088] [Medline: 30401667]

39. Mori M, MacDorman K, Kageki N. The uncanny valley [from the field]. IEEE Robot Automat Mag 2012 Jun;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]

40. MacDorman KF, Green RD, Ho C, Koch CT. Too real for comfort? Uncanny responses to computer generated faces. Comput Human Behav 2009 May 01;25(3):695-710 [FREE Full text] [doi: 10.1016/j.chb.2008.12.026] [Medline: 25506126]

41. Cheetham M, Pavlovic I, Jordan N, Suter P, Jancke L. Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. Front Psychol 2013;4:1-12 [FREE Full text] [doi: 10.3389/fpsyg.2013.00108] [Medline: 23471214]

42. Lay S, Brace N, Pike G, Pollick F. Circling around the uncanny valley: design principles for research into the relation between human likeness and eeriness. Iperception 2016;7(6):1 [FREE Full text] [doi: 10.1177/2041669516681309] [Medline: 27994844]

43. Thompson J. EHC Interactive Clinical Avatar Demonstration. 2019. URL: https://youtu.be/wMx9upGgUcE [accessed 2020-10-23]

44. Thompson J, White S, Chapman S. Actual vs. perceived competency development: how can virtual patients impact pharmacist pre-registration training? Pharmacy (Basel) 2020 Aug 05;8(138):1 [FREE Full text] [doi: 10.3390/pharmacy8030138] [Medline: 32764475]

45. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]

46. Morse JM. The significance of saturation. Qual Health Res 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]

47. Salter SM, Karia A, Sanfilippo FM, Clifford RM. Effectiveness of E-learning in pharmacy education. Am J Pharm Educ 2014 May 15;78(4):16-22 [FREE Full text] [doi: 10.5688/ajpe78483] [Medline: 24850945]

48. Watson M, Norris P, Granas A. A systematic review of the use of simulated patients and pharmacy practice research. Int J Pharm Pract 2006;14(2):83-93. [doi: 10.1211/ijpp.14.2.0002]

49. Kardong-Edgren S, Adamson KA, Fitzgerald C. A review of currently published evaluation instruments for human patient simulation. Clin Simul Nurs 2010 Jan;6(1):e25-e35. [doi: 10.1016/j.ecns.2009.08.004]

50. Abdo A, Ravert P. Student satisfaction with simulation experiences. Clin Simul Nurs 2006 Jan;2(1):e13-e16. [doi: 10.1016/j.ecns.2009.05.009]

51. De Leng B, Huwendiek S, Donkers J, eVIP Project Team. Published evaluation instrument for virtual patients and for learning and teaching activities with virtual patients. 2009 May 15. URL: https://www.virtualpatients.eu/wp-content/uploads/2009/06/evip_deliverable_5.2_final-3.pdf [accessed 2020-10-23]

52. Lambton J, O'Neill SP, Dudum T. Simulation as a strategy to teach clinical pediatrics within a nursing curriculum. Clin Simul Nurs 2008 Oct;4(3):e79-e87. [doi: 10.1016/j.ecns.2008.08.001]

53. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]

54. Johnson RB, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. Educ Res 2004 Jul;33(7):14-26. [doi: 10.3102/0013189X033007014]

55. Frey B, Edwards L. Strong words or moderate words: a comparison of the reliability and validity of responses on attitude scales. Psychology 2011;2(1):49-52. [doi: 10.4236/psych.2011.21008]

56. Wilkinson S. Women with breast cancer talking causes: comparing content, biographical and discursive analyses. Fem Psychol 2000;10(4):431-460. [doi: 10.1177/0959353500010004003]

XSL·FO

RenderX

57. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]

58. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]

59. Baumann-Birkbeck L, Florentina F, Karatas O, Sun J, Tang T, Thaung V, et al. Appraising the role of the virtual patient for therapeutics health education. Curr Pharm Teach Learn 2017 Dec;9(5):934-944. [doi: 10.1016/j.cptl.2017.05.012] [Medline: 29233327]

60. Smith MA, Benedict N. Effectiveness of educational technology to improve patient care in pharmacy curricula. Am J Pharm Educ 2015 Feb 17;79(1):15 [FREE Full text] [doi: 10.5688/ajpe79115] [Medline: 25741031]

61. Zary N, Johnson G, Boberg J, Fors UGH. Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment (Web-SP). BMC Med Educ 2006 Feb 21;6:10 [FREE Full text] [doi: 10.1186/1472-6920-6-10] [Medline: 16504041]

62. Tait L, Lee K, Rasiah R, Cooper JM, Ling T, Geelan B, et al. Simulation and feedback in health education: a mixed methods study comparing three simulation modalities. Pharmacy (Basel) 2018 May 03;6(2):1 [FREE Full text] [doi: 10.3390/pharmacy6020041] [Medline: 29751528]

63. Smith MA, Mohammad RA, Benedict N. Use of virtual patients in an advanced therapeutics pharmacy course to promote active, patient-centered learning. Am J Pharm Educ 2014 Aug 15;78(6):125 [FREE Full text] [doi: 10.5688/ajpe786125] [Medline: 25147397]

64. Menendez E, Balisa-Rocha B, Jabbur-Lopes M, Costa W, Nascimento JR, Dósea M, et al. Using a virtual patient system for the teaching of pharmaceutical care. Int J Med Inform 2015 Sep;84(9):640-646. [doi: 10.1016/j.ijmedinf.2015.05.015] [Medline: 26077512]

65. Barnett SG, Gallimore CE, Pitterle M, Morrill J. Impact of a paper vs virtual simulated patient case on student-perceived confidence and engagement. Am J Pharm Educ 2016 Feb 25;80(1):9-14 [FREE Full text] [doi: 10.5688/ajpe80116] [Medline: 26941442]

66. Fry A, Ketteridge S, Marshall S. Understanding student learning. In: Fry H, Ketteridge S, Marshall S, editors. A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice. 2nd Edition. Philadelphia: Routledge; 2009:9-25.

67. Prensky M. Digital natives, digital immigrants. Part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]

68. Herrington J, Oliver R, Reeves T. "Cognitive realism" in online authentic learning environments. 2003 Presented at: Association for the Advancement of Computing in Education; 2003; Honolulu p. 2115-2121. [doi: 10.1007/978-1-4614-3185-5_32]

69. Kolb AY, Kolb DA. Learning styles and learning spaces: enhancing experiential learning in higher education. Acad Manag Learn Educ 2005 Jun;4(2):193-212. [doi: 10.5465/amle.2005.17268566]

70. Salinitri FD, O'Connell MB, Garwood CL, Lehr VT, Abdallah K. An objective structured clinical examination to assess problem-based learning. Am J Pharm Educ 2012 Apr 10;76(3):44 [FREE Full text] [doi: 10.5688/ajpe76344] [Medline: 22544961]

71. Forsberg E, Georg C, Ziegert K, Fors U. Virtual patients for assessment of clinical reasoning in nursing: a pilot study. Nurse Educ Today 2011 Nov;31(8):757-762. [doi: 10.1016/j.nedt.2010.11.015] [Medline: 21159412]

72. Bateman J, Allen ME, Kidd J, Parsons N, Davies D. Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study. BMC Med Educ 2012 Aug 01;12:62 [FREE Full text] [doi: 10.1186/1472-6920-12-62] [Medline: 22853706]

73. Poulton T, Balasubramaniam C. Virtual patients: a year of change. Med Teach 2011 Jan;33(11):933-937. [doi: 10.3109/0142159X.2011.613501] [Medline: 22022903]

74. Al-Dahir S, Bryant K, Kennedy KB, Robinson DS. Online virtual-patient cases versus traditional problem-based learning in advanced pharmacy practice experiences. Am J Pharm Educ 2014 May 15;78(4):1-8 [FREE Full text] [doi: 10.5688/ajpe78476] [Medline: 24850938]

75. Battaglia JN, Kieser MA, Bruskiewitz RH, Pitterle ME, Thorpe JM. An online virtual-patient program to teach pharmacists and pharmacy students how to provide diabetes-specific medication therapy management. Am J Pharm Educ 2012 Sep 10;76(7):131 [FREE Full text] [doi: 10.5688/ajpe767131] [Medline: 23049103]

76. Fuhrman J, Buff W, Esddy M, Dollar M. Utilization of an integrated interactive virtual patient database in a web-based environment for teaching continuity of care. Am J Pharm Educ 2001;65(3):271-275 [FREE Full text] [doi: 10.5688/aj650313]

77. Olin JL, Cole SW. Computer-based clinical simulation cases in oncology pharmacotherapy: student perceptions from two campuses. Curr Pharm Teach Learn 2015 Mar;7(2):157-162. [doi: 10.1016/j.cptl.2014.11.012]

78. Orr KK. Integrating virtual patients into a self-care course. Am J Pharm Educ 2007 Apr 15;71(2):1-9 [FREE Full text] [doi: 10.5688/aj710230] [Medline: 17533439]

79. Smith MA, Siemianowski LA, Benedict N. Virtual patient case sharing across two schools of pharmacy. Am J Pharm Educ 2016 Nov 25;80(9):1 [FREE Full text] [doi: 10.5688/ajpe809153] [Medline: 28090102]

80. Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach 2008;30(2):170-174. [doi: 10.1080/01421590701874074] [Medline: 18464142]

81. Harden RM. What is a spiral curriculum? Med Teach 1999;21(2):141-143. [doi: 10.1080/01421599979752] [Medline: 21275727]
82. Krathwohl D. A revision of Bloom's taxonomy: an overview. Theory Pract 2002;41:212-218. [doi: 10.1207/s15430421tip4104]
83. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010 Jan;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]

Abbreviations

3D: 3-dimensional
EHC: emergency hormonal contraception
OSCE: Objective Structured Clinical Examination

XSL•FO
RenderX

Original Paper

The Role of Enjoyment in a Serious Game for Binge Drinking Prevention: Pretest-Posttest Quasi-Experimental Study

Traci Hong[1*], PhD; Joshua Cabrera[2*], MD; Christopher E Beaudoin[1*], PhD

[1]College of Communication, Boston University, Boston, MA, United States

[2]MHMR Authority of Brazos Valley, Bryan, TX, United States

[*]all authors contributed equally

Corresponding Author:
Traci Hong, PhD
College of Communication
Boston University
640 Commonwealth Ave
Boston, MA, 02215
United States
Phone: 1 6173533450
Email: tjhong@bu.edu

Abstract

Background: Although binge drinking peaks at age 21 to 25 years, there is limited research on the effects of serious games in this population, as well as on the process by which playing serious games impacts alcohol-related outcomes. Designed with both health behavioral theory and game theory, One Shot is an online serious game that aims to prevent binge drinking.

Objective: This study utilized a conceptual model for serious video game processes. Using One Shot, the model assessed the following process stages: (1) Alcohol Use Disorders Identification Test-Concise (AUDIT-C); (2) in-game factors of game time and risky alcohol decisions; (3) game enjoyment; and (4) postgame outcomes of intention to drink less and drinking refusal self-efficacy.

Methods: In a one-group pretest-posttest quasi-experimental design, a sample (N=550) of young adults (age 21-25 years) who reported recent binge drinking played the One Shot game. Intention to drink less and drinking refusal self-efficacy were measured at pregame and postgame, with their effects lagged in statistical analysis. Participants were presented with various scenarios in the game that pertained to risky alcohol decisions, which, along with game time, were unobtrusively recorded by the server. A structural equation model (SEM) was used to test the conceptual model, with assessments made to determine if enjoyment mediated the effects of game time and risky alcohol decisions on the 2 postgame alcohol-related outcomes.

Results: A well-fitting SEM demonstrated support for the multistep model, with AUDIT-C predicting risky alcohol decisions (β=.30). Risky alcohol decisions (β=−.22) and game time (β=.18) predicted enjoyment, which, in turn, predicted intention to drink less (β=.21) and drinking refusal self-efficacy (β=.16). Enjoyment significantly (P<.001) mediated the effects of game time and risky alcohol decision on intention to drink less and drinking refusal self-efficacy.

Conclusions: The results support a conceptual model in which staggered individual and in-game factors influence alcohol-related outcomes. Enjoyment is important for participants' intentions to drink less and beliefs that they can refuse alcohol.

Introduction

Background

Although it is well recognized that alcohol use is a leading risk factor for morbidity and mortality, recent estimates indicate that it accounts for a larger percentage of global deaths than previously recognized, with 10% of deaths among 15 to 49-year-old individuals attributed to alcohol [1]. Excessive alcohol use is associated with considerable health burdens and outcomes, including cancer, heart disease, and stroke [2-4]. Binge drinking, which is defined as consuming 5 or more drinks for men and 4 or more drinks for women in a 2-hour period [5],

XSL·FO
RenderX

is common, with more than a quarter of US adults reporting it [6]. To prevent alcohol misuse, video games have been used to impart knowledge and change health behaviors [7], with much of this research focused on adolescents and college students because of the health and injury risks associated with underage drinking [8]. Designed to entertain, as well as simultaneously impart knowledge and model health behaviors [9,10], these "serious video games" are interactive interfaces with a single player or multiple players on different platforms, including online, computer, game console, and mobile [10].

While there are several serious games on substance use, few are exclusively focused on alcohol, with nearly all such games targeting adolescents and college students [8,11-13]. Research on serious games has demonstrated participants' development of knowledge on harms and attitudes about alcohol consumption [8]. In terms of behavioral change, there is evidence that serious games can result in increased drinking refusal skills [8] and short-term effects on reduced alcohol consumption [13], particularly among females [12] and younger adolescents [12]. In contrast, some studies found inconclusive effects [14] or no effect [11].

There are very few studies specific to adults of legal drinking age (ie, 21 years or older) in the United States. Though understudied, young adults (ie, from 21 to 25 years of age) represent an important population for alcohol-prevention gaming because excessive alcohol consumption peaks at the end of or even after college, with binge drinking levels at a high for women at age 22 years and for men at age 23 years [15]. These peak ages for binge drinking reflect a steady increase across the last 30 years, whereby the peak age of binge drinking has increased by 2 years [15]. In addition, the highest weekly alcohol intake occurs at about age 25 years for both men and women [16]. Young adults are also an important group for alcohol-prevention video games given that the majority of people in this age group have played video games [17].

The aim of this study was to explain how individual characteristics and in-game factors of a serious game can influence drinking-related outcomes for young adults aged 21 to 25 years. To study the underlying mechanisms, we build a multistep model that entails individual factors, in-game factors, and postgame alcohol outcomes. Below, we describe our serious video game and its theoretical basis. Then, we articulate expectations related to our conceptual model.

"One Shot" Online Video Game

This study's serious game "One Shot" is an online single-player game designed by a physician (ie, this study's second author) specializing in alcohol addiction among young adults. Single-player games can be particularly adept at imparting new skills and knowledge [18]. One Shot is a 2-dimensional web-based game created on the Unity Game platform with a role-playing story-based design in which players make decisions that influence their relationships with other game characters. The game characters are nonplayer characters who the game player does not control. Subtle background noise (eg, chatter of people at a party) and music appear throughout the game. The dialogue and narrative description are displayed in a box at the bottom of the screen. Player interaction is facilitated with mouse clicks or keyed selections. The player advances through the narrative description and makes decisions by selecting the preferred option. The dialogue was pilot tested with 21-year-old individuals. The game can be run on the Chrome or Firefox browser. Figure 1 depicts the scene game players encounter upon arriving at a party.

Figure 1. Screenshot of the party scene of the One Shot game.
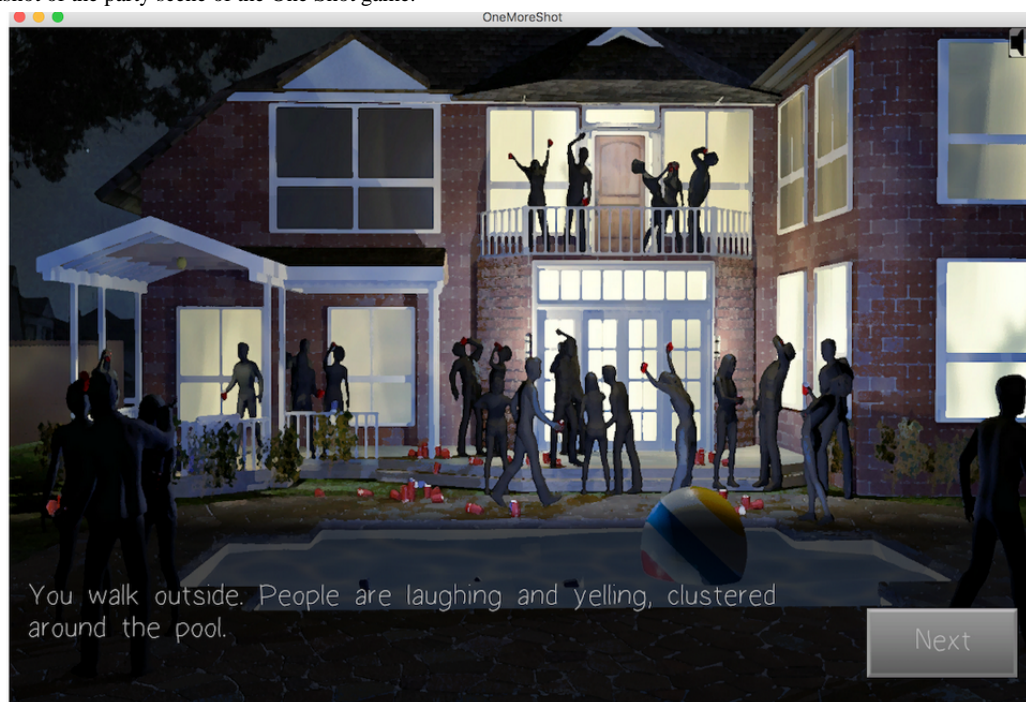
Theoretical Basis

The design of One Shot utilizes both game theory and health behavioral theory given that the combination of both theoretical approaches elicits the strongest health outcomes [19]. In terms of game theory, central to an effective serious game is a story-telling element [9,10]. In particular, role-playing games are adept at facilitating a video game player's identification with characters and roles [20]. This type of identification can subsequently lead to game enjoyment [20], which is important to alcohol-related outcomes. The video game narrative begins when two close friend nonplayer characters arrive to pick up the game player to go to an evening party at a house where alcohol is widely consumed by guests. Conflict emerges when the game player is exposed to the opposing perspectives on alcohol consumption of the two friends and needs to choose between varying levels of risky alcohol consumption. Figure 2 depicts one of the close friend nonplayer characters at the party and an example of a high-risk alcohol consumption opportunity (ie, consumption of strong alcohol) that is presented to game players. Irrespective of the risk decisions made in One Shot, all game participants experience a car accident involving a physical injury and visit to the emergency room, thus highlighting the severe negative consequences of drinking. The game player and close friend nonplayer characters meet a nonplayer character nurse in the emergency room, who explains that a car accident occurred and then educates them about the hazards of binge drinking.

In terms of health behavioral theory, protection motivation theory (PMT) has the following 2 core concepts: perceived threat and coping. It suggests people are likely to take action to avert a severe threat if they are provided with coping strategies [21]. In the emergency room scene of One Shot, coping strategies are conveyed by tailored feedback that is given to game players after they are asked about their alcohol-related knowledge and perceptions. For example, Figure 3 depicts tailored feedback provided to game players who do not understand the size of standard alcoholic drinks. Other instances of tailored feedback entail specific coping strategies to avert binge drinking.

In addition, given that previous research has demonstrated that social norms influence alcohol consumption [22,23], One Shot aims to utilize the normative influence of close friend nonplayer characters. The game provides a mix of peer approval and disapproval of alcohol consumption. In particular, there is strong disapproval from the nurse, whereas, of the two friends at the beginning of the game, one strongly approves of binge drinking and the other disapproves of binge drinking. Across the game, the nonplayer character friend who was initially prodrinking changes to disapprove of binge drinking. With this transition, both nonplayer character friends, as well as the nonplayer character nurse, strongly disapprove of binge drinking by the end of the game.

Figure 2. Screenshot of a close friend nonplayer character encouraging the game player to drink strong alcohol.

Figure 3. Screenshot of tailored feedback for game players.
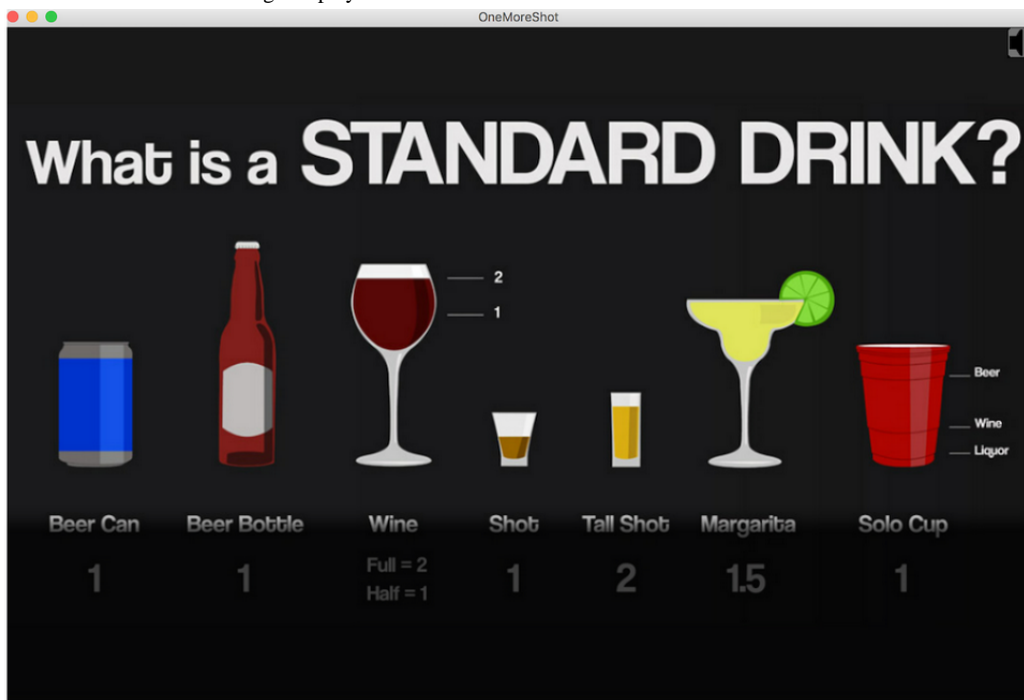
A Conceptual Model for Serious Video Game Processes

Our conceptual model for serious video game processes is depicted in Figure 4. It entails the following 3 main stages of factors: (1) individual factors; (2) in-game factors; and (3) postgame outcomes. In-game factors have 2 substages entailing time spent playing the game and risky alcohol decisions that are made (both of which are unobtrusively recorded by the host server) and enjoyment of the game (which is measured immediately after the game).

Figure 4. Conceptual model for serious video game processes.
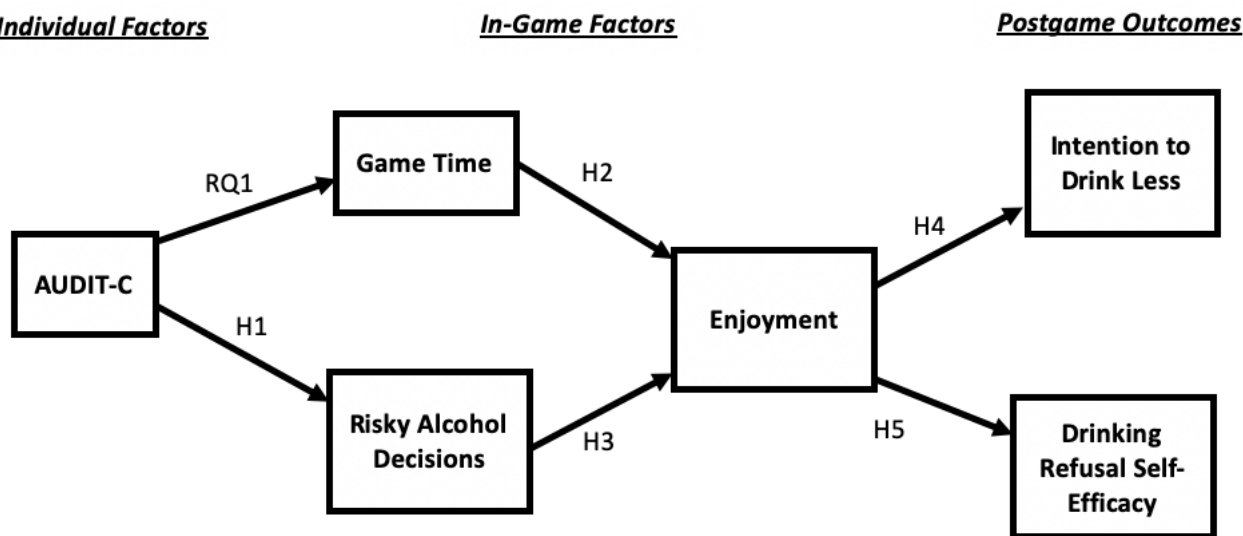
Individual Factors

The first set of factors in our conceptual model is the individual characteristics of the game player. Prior research has implemented health predispositions as an individual factor [24]. For example, AUDIT-C was implemented as an individual factor for a serious game that assessed alcohol outcomes and participants' willingness to seek help for health problems [13]. AUDIT-C measures risk for hazardous alcohol consumption and alcohol use disorder [25]. These studies suggest that measures of health disorders, such as AUDIT-C, are individual factors that can affect how people play serious video games and are influenced by them.

In-Game Factors and Their Antecedents

The second main stage in our conceptual model entails in-game factors specific to the One Shot serious game. In Figure 4, they appear in the following 2 substages: (1) game time and risky alcohol decisions and (2) enjoyment. In the first substage, game time entails how long a player was active in the game environment, while risky alcohol decisions involve how many risky choices game players made when presented with various

scenarios that entailed alcohol consumption, including low-risk decisions such as drinking one alcoholic beverage, but refusing another, and high-risk decisions such as elevating drinking to hard liquor, playing a drinking game, and drinking and driving.

Pertinent to this stage in the model, research indicates that video game players behave in a virtual environment similarly to how they behave in the real world [26]. Thus, we expect that individuals who exhibit hazardous drinking in real life will exhibit similar drinking behaviors in the game. Given that AUDIT-C is a reliable measure for alcohol abuse disorder and hazardous drinking [27], which is associated with risky decision making [28], we expect that game players who are more likely to exhibit hazardous drinking (ie, higher levels of AUDIT-C) will choose to partake in more risky alcohol behaviors when they are presented with such options in the video game. The first hypothesis (H1) is as follows: AUDIT-C predicts an increase in risky alcohol decisions.

Unclear, however, is the effect of hazardous drinking on time spent playing the game. The first research question (RQ1) is as follows: What is the effect of AUDIT-C on game time?

The second substage of in-game factors (Figure 4) entails enjoyment of the game. Important to media entertainment on any platform (eg, video game, film, and television) is a user's sense of enjoyment, which is defined as a positive reaction to media content [29], encompassing both cognitive and affective dimensions [29]. A fundamental aspect of serious games is that learning and health behavior change occur alongside entertainment [9,10], that is, video game players are entertained while, through social modeling [30], they acquire knowledge, which can lead to behavioral adoption.

Despite the suggested importance of enjoyment, few serious games specific to alcohol prevention account for the role of enjoyment in influencing related game outcomes. In our conceptual model, the in-game factors of game time and risky alcohol decisions predict game enjoyment. Previous research found that in-game experiences are associated with enjoyment [31]. Additionally, previous research documented that the more time people spend playing a video game, the more likely they are to report enjoying the game [32]. The second hypothesis (H2) is as follows: Game time predicts an increase in enjoyment.

Next, we considered the effects of risky alcohol decisions. In the One Shot game, players who make more risky alcohol decisions are more extensively exposed to game scenarios that lead to the game's severe consequences (eg, car accident and visit to a hospital). These severe consequences serve as fear appeals, which elicit an unpleasant emotional state that arises from a threat [33], and in PMT, result from perceptions of the severity and susceptibility of the threat [21]. In the video game, the severity of the threat encompasses the car accident and emergency room visit, whereas the susceptibility of the threat entails the escalating storylines as players make more risky alcohol decisions. Making more risky decisions would thus likely elicit negative emotions and, in turn, lower levels of game enjoyment. The third hypothesis (H3) is as follows: Risky alcohol decisions predict a decrease in enjoyment.

Postgame Outcomes and Their Antecedents

The third stage in our conceptual model (Figure 4) entails the following 2 postgame outcomes: intention to drink less and drinking refusal self-efficacy. Behavioral intention is defined as how likely a person thinks he/she would perform a specific behavior and is highly predictive of a person actually performing such a behavior in the future [34]. Drinking refusal self-efficacy is conceptualized as one's belief in his/her ability to refuse solicitation of alcohol in commonly encountered situations (eg, when a friend dares you to drink in excess) [35] and is highly predictive of alcohol consumption frequency and volume [36].

We consider the effects of enjoyment on both of these outcomes. Enjoyment of a game underpins the entertainment media experience, such that attaining game enjoyment is necessary for substantive effects, most notably learning outcomes [29]. Previous research on serious games has documented the importance of game enjoyment, with people's enjoyment of a game being correlated with their positive attitudes toward the game [37], as well as their self-efficacy [38] and intention to perform a recommended behavior [39]. Given that we lag the postgame outcomes by analogous measures at pregame, our hypotheses pertain to the change in outcomes. The fourth hypothesis (H4) is as follows: Enjoyment predicts postgame intention to drink less (when controlling for its pregame level). The fifth hypothesis (H5) is as follows: Enjoyment predicts postgame drinking refusal self-efficacy (when controlling for its pregame level).

Given the multistage aspect of our conceptual model for serious video game processes (Figure 4), we also considered the mediation role played by enjoyment. Because enjoyment is a central concept in entertainment media [29] and entertainment serves as the foundation for learning in serious games [9,10], it is expected that enjoyment plays an integral role in the One Shot game, mediating the effects of prior in-game factors on postgame outcomes. The sixth hypothesis (H6) is as follows: Enjoyment mediates the effects of game time on postgame intention to drink less and drinking refusal self-efficacy (when controlling for the pregame levels). The seventh hypothesis (H7) is as follows: Enjoyment mediates the effects of risky alcohol decisions on postgame intention to drink less and drinking refusal self-efficacy (when controlling for their pregame levels).

Methods

Sample and Design

This study utilized a one-group pretest-posttest quasi-experimental design [40], which has been used to evaluate serious video games for alcohol prevention [13], as well as other health education topics [41,42]. The One Shot video game was implemented online and made accessible to study participants. There were online survey assessments before (ie, pregame) and after (ie, postgame) the online video game. Measures for 2 in-game factors (ie, game time and risky alcohol decisions) were captured through log files on the host server. The sample of participants was recruited by Survey Sampling International (SSI) and is of US young adults aged 21 to 25 years who reported binge drinking in the 2 weeks preceding this study

XSL•FO

RenderX

(N=550). SSI drew participants from a nationally representative internet panel of individuals from all 50 states, and the sample is representative of national demographic parameters. To build the panel, the SSI panel used targeted approaches, including search links, banner advertisements, email, online invitations, and coregistration. SSI also employed data-validation techniques, such as comparing participant demographics with multiple databases and data vendors, to verify personal identifying information. The pregame and postgame questionnaires were conducted via Qualtrics in June and July 2016. Institutional Review Board approval for this study was acquired at Texas A&M University.

Measurements

There were the following 4 demographics: age, education, ethnicity (W=1), and gender (M=1). Education was measured on a 7-point scale from less than high school graduate (1) to doctoral degree (7) and recoded to represent years of study. Other variables were AUDIT-C, game time, risky alcohol decisions, enjoyment, intention to drink less, and drinking refusal self-efficacy. AUDIT-C was assessed at pregame, whereas enjoyment was assessed at postgame. Intention to drink less and drinking refusal self-efficacy were assessed at pregame and postgame.

With scores ranging from 0 to 12, AUDIT-C entails 3 survey items that help identify people who are hazardous drinkers or have active alcohol-use disorders [25]. Game time was measured in the following 4 ordered groups: (1) 0-500 seconds; (2) 501-1000 seconds; (3) 1001-1500 seconds; and (4) 1500-1800 seconds. Risky alcohol decisions involved the following 2 items: low-risk alcohol decisions and high-risk alcohol decisions. There were 2 survey items for enjoyment [43]. The first question entailed how much participants enjoyed the game experience, with responses on a 5-point scale from "I really didn't enjoy it" (1) to "I really enjoyed it" (5). The second question entailed the likelihood of participants recommending the game to a friend, with responses on a 5-point scale from "strongly recommend" (1) to "strongly not recommend" (5). This second set of responses was reversed coded, so that higher responses to both questions represented more positive impressions of the video game (r=0.54, $P<.001$). Intention to drink less was measured with one item on a 5-point scale (with responses from 0-4), involving whether participants planned to moderate their alcohol use in the future [44]. Drinking refusal self-efficacy was assessed with 7 items [45]. For example, participants were asked how sure they were that they could stop drinking if there were problems with friends or for themselves, with responses on a 5-point scale from "not at all sure" (1) to "extremely sure" (5). This scale was internally consistent at pregame (α=.83) and postgame (α=.88).

Statistical Analysis

Stata 16 (Stata Corp) was used for statistical analysis. No outliers were identified. Square root transformations were conducted for skewed-right variables (ie, AUDIT-C, game time, and the 2 manifest indicators of risky alcohol decisions). Square transformations were implemented for skewed-left variables (ie, intention to drink less and the 2 manifest indicators of enjoyment). Descriptive statistics, as well as t tests comparing

pregame and postgame means of the dependent variables, were used for variables in their pretransformation form. Posttransformation variables were used for the structural equation model (SEM), which was implemented using the maximum likelihood method of estimation and covariance structure analysis. SEM is a technique for confirming a specified model of multivariate relationships [46]. Entailing simultaneous regressions, SEM permits the testing of the stage-like effects of our conceptual model for serious video game processes. The excellence of model fit is assessed according to the benchmarks of Hu and Bentler [47] as follows: a comparative fit index (CFI) value ≥0.95 and a root mean square error of approximation (RMSEA) value of close to 0.06 or less.

Exogenous control variables included age, education, ethnicity (W=1), and gender (M=1), as well as the pregame measures of intention to drink less and drinking refusal self-efficacy. Endogenous variables included AUDIT-C, game time, risky alcohol decisions, and enjoyment, as well as postgame measures of intention to drink less and drinking refusal self-efficacy. Paths were drawn from exogenous variables to endogenous variables and among endogenous variables, as depicted in Figure 4. Using an autoregression approach, paths from pregame to postgame intention to drink less and from pregame to postgame drinking refusal self-efficacy entail cross-lagged effects, which represent the stability of a measure over time and control the postgame variable for the pregame level of the variable [48]. In this manner, the effects of the other exogenous and endogenous variables on postgame intention to drink less and postgame drinking refusal self-efficacy are indicative of their influence on the changes from pregame to postgame in these health outcomes. There were the following 2 covariance paths: between game time and risky alcohol decisions and between postgame intention to drink less and postgame drinking refusal self-efficacy.

In the SEM, enjoyment, risky alcohol decisions, and pregame and postgame drinking refusal self-efficacy were instituted as latent variables, while the other variables were observed. In the first step of the SEM approach, confirmatory factor analysis (CFA) was conducted to test the unidimensionality of the latent variables on pregame and postgame drinking refusal self-efficacy. Second, a structural model was tested to assess relationships among the latent and observed variables. Mediation was assessed with the product of coefficients approach [49].

Results

Descriptive statistics are depicted in Table 1. In their pretransformation form, the mean for intention to drink less was significantly higher at postgame than at pregame (1.61 vs 1.46; t_{549}=−2.97, P=.003). Similar tests were run on the 7 pregame and postgame manifest indicators of drinking refusal self-efficacy, which was instituted as a latent variable in the SEM. In the following 5 cases, the item means were higher at postgame than at pregame (Table 1): (1) item 1, t_{549}=−5.51, $P<.001$; (2) item 3, t_{549}=−5.29, $P<.001$; (3) item 4, t_{549}=−5.89, $P<.001$; (4) item 5, t_{549}=−8.11, $P<.001$; and (5) item 6, t_{549}=−2.65, P=.008. In the other 2 cases, the item means did not vary significantly from pregame to postgame (Table 1) as

follows: item 2, $t_{549}=-1.93$, $P=.054$ and item 7, $t_{549}=-0.57$, $P=.57$.

Table 1. Descriptive statistics (N=550).

Variable	Value
Age (years), mean (SD)	22.65 (1.18)
Education (years), mean (SD)	14.41 (2.10)
AUDIT-C[a] (0-12 index), mean (SD)	4.11 (2.17)
Game time (1-4 scale), mean (SD)	1.81 (2.17)
Low-risk decisions (number), mean (SD)	1.61 (1.38)
High-risk decisions (number), mean (SD)	3.56 (3.30)
Game enjoyment (1-5 scale), mean (SD)	3.87 (1.13)
Game recommendation (1-5 scale), mean (SD)	4.13 (1.01)
Pregame intention to drink less (0-4 scale), mean (SD)	1.46 (1.36)
Postgame intention to drink less (0-4 scale), mean (SD)	1.61 (1.36)
Pregame drinking refusal self-efficacy, mean (SD)	
Item 1 (1-5 scale)	3.58 (1.26)
Item 2 (1-5 scale)	4.09 (1.21)
Item 3 (1-5 scale)	3.55 (1.27)
Item 4 (1-5 scale)	3.30 (1.36)
Item 5 (1-5 scale)	3.01 (1.38)
Item 6 (1-5 scale)	3.89 (1.30)
Item 7 (1-5 scale)	4.05 (1.22)
Postgame drinking refusal self-efficacy, mean (SD)	
Item 1 (1-5 scale)	3.86 (1.29)
Item 2 (1-5 scale)	4.18 (1.18)
Item 3 (1-5 scale)	3.81 (1.23)
Item 4 (1-5 scale)	3.63 (1.33)
Item 5 (1-5 scale)	3.49 (1.37)
Item 6 (1-5 scale)	4.02 (1.21)
Item 7 (1-5 scale)	4.07 (1.22)
Ethnicity, n (%)	
Other	229 (41.64)
White	321 (58.36)
Gender, n (%)	
Female	317 (57.64)
Male	233 (42.36)

[a]AUDIT-C: Alcohol Use Disorders Identification Test-Concise.

CFA included the latent structures for pregame and postgame drinking refusal self-efficacy. The initial CFA included covariance paths between pregame and postgame drinking refusal self-efficacy and between each pregame manifest indicator and its analogous postgame manifest indicator. The tested CFA had decent model fit (CFI=0.944; RMSEA=0.080; $\chi^2_{69}=312.61$, $P<.001$). According to modification indices, 2 additional covariance paths were added between latent variable indicators. The retested CFA had good model fit (CFI=0.977; RMSEA=0.052; $\chi^2_{67}=167.65$, $P<.001$). This latent variable structure, with factor loadings on the 2 factors, was used in the structural model, which was calculated next. The tested SEM had good model fit (CFI=0.960; RMSEA=0.039; $\chi^2_{246}=449.08$, $P<.001$). The Bentler-Raykov squared multiple correlation coefficients were as follows: AUDIT-C, 5.58%; game time, 2.01%; risky alcohol decisions, 11.53%; enjoyment, 17.22%; postgame intention to drink less, 40.57%; and postgame drinking refusal self-efficacy, 70.35%.

The effects of variables are depicted in Figure 5. Both lagged effects were significant (intention to drink less, β=.60, *P*<.001; drinking refusal self-efficacy, β=.77, *P*<.001), which suggests over-time stability in measurement. The positive effect of AUDIT-C on risky alcohol decisions was significant (β=.30, *P*<.001), supporting H1. In terms of RQ1, the effect of AUDIT-C on game time was not significant (β=−.03). The positive effect of game time on enjoyment was significant

(β=.18, *P*<.001), supporting H2. The negative effect of risky alcohol decisions on game enjoyment was significant (β=−.22, *P*<.001), supporting H3. The positive effect of enjoyment on postgame intention to drink less was significant (β=.21, *P*<.001), supporting H4. The positive effect of game enjoyment on postgame drinking refusal self-efficacy was significant (β=.16, *P*<.001), supporting H5.

Figure 5. Direct effects of variables in the structural equation model (N=550). *P<.05, **P<.01, ***P<.001.
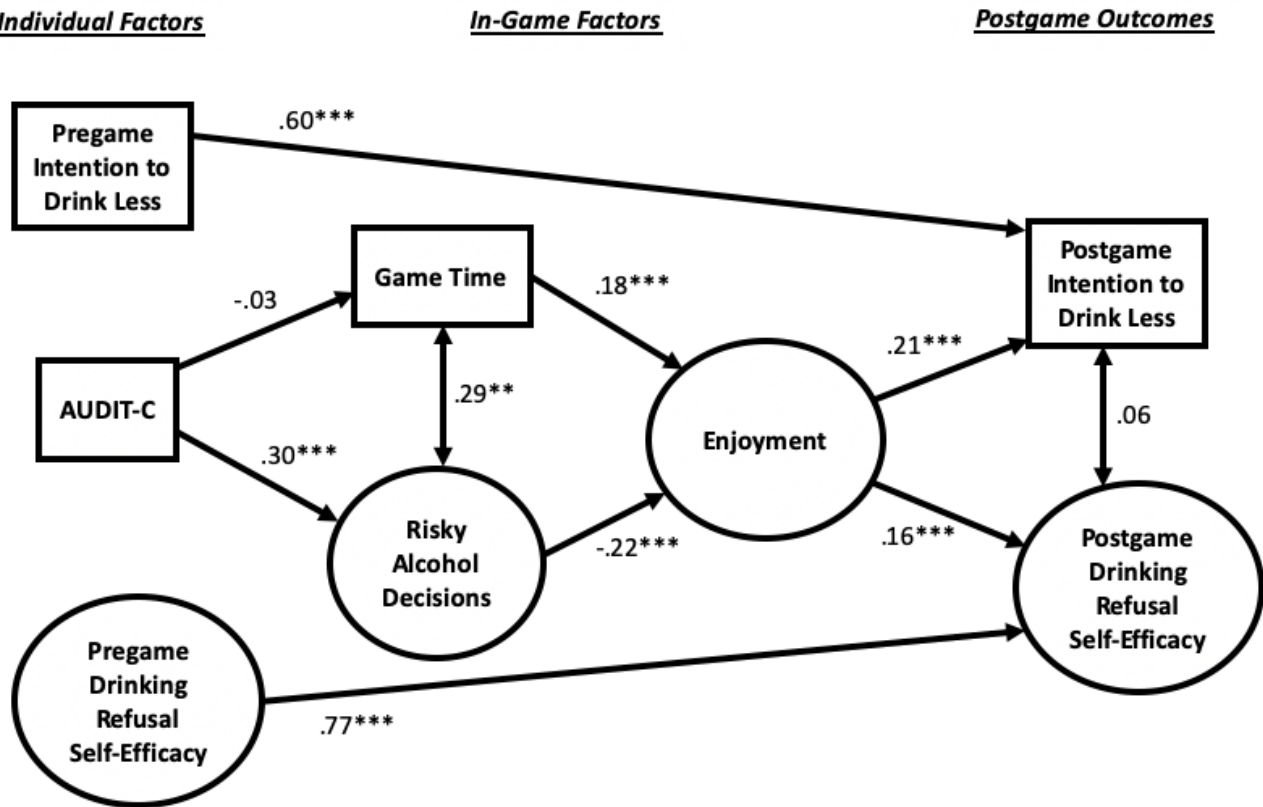
H6 and H7 entail mediation paths. The significant role of enjoyment in mediating the effect of game time on intention to drink less was documented, with a product of coefficients of 13.28 (*P*<.001) [50]. Similarly, the significant role of enjoyment in mediating the effect of game time on drinking refusal self-efficacy was demonstrated, with a product of coefficients of 11.98 (*P*<.001) [50]. These findings support H6. The significant role of enjoyment in mediating the effect of risky alcohol decisions on intention to drink less was demonstrated, with a product of coefficients of −7.78 (*P*<.001) [50]. Similarly, the significant role of enjoyment in mediating the effect of risky alcohol decisions on drinking refusal self-efficacy was documented, with a product of coefficients of 7.02 (*P*<.001) [50]. These findings support H7.

Discussion

Principal Findings

Our conceptual process model depicts how a serious game for alcohol prevention influences changes in alcohol-related outcomes. With our sample of young adults aged 21 to 25 years, which is an understudied yet at-risk group for binge drinking, the analysis suggested improvements from pregame to postgame

in intention to drink less, as well as in 5 of the 7 manifest indicators of drinking refusal self-efficacy. The SEM approach supports the multistep conceptual model, moving from individual factors to 2 staggered substages of in-game factors to postgame alcohol outcomes. In a practical sense, the effects on intention to drink less and drinking refusal self-efficacy are important given their likelihood of predicting subsequent decreases in alcohol consumption behavior [34,36]. Of theoretical importance, our process model contains conceptual components from both health behavioral and game theories given that research has demonstrated that this joint utilization enhances the likelihood that serious games achieve their desired outcomes [19]. Importantly, the results pertain to an at-risk population as all study participants reported binge drinking during the prior 2 weeks.

Consistent with game theory, the extent that game players enjoyed One Shot elicited increases from pregame to postgame in their intention to drink less and drinking refusal self-efficacy. Notably, enjoyment mediated the effects of the following 2 in-game factors: game time and risky alcohol decisions. Thus, instead of these in-game factors having direct effects on alcohol outcomes, their documented influence was indirect, as mediated

by enjoyment. These indirect effects are suggestive of the complex processes by which an alcohol-prevention serious game functions, with the concept of enjoyment at its core, which is consistent with media entertainment theory that posits that enjoyment is central to learning-related outcomes for entertainment media in general and serious games in particular [29].

These results on the mediating role of enjoyment suggest the need to integrate entertainment content, components, and appeals in serious game design and development as a means to foster learning and behavioral outcomes related to health. The combination of behavioral theory and game theory is needed to elicit desired health-related outcomes [19]. The design of our game is based on social cognitive theory [30], whereby game players learn through the social modeling of nonplayer characters, and PMT, which emphasizes the importance of threat and coping [21]. In addition, game theory specifies that a storytelling format is required to engage players in a serious game [9,10]. Important here is that game players have a degree of control in their decision making and subsequent trajectory across the game experience, which likely contributes to their level of game enjoyment.

It is also important to consider the role of AUDIT-C in the process. Participants with higher scores on AUDIT-C selected more in-game risky alcohol decisions, which is emblematic of how in-game decisions mirror real-world behaviors. In turn, participants who made more risky decisions enjoyed the game less. In terms of the overall process, participants with higher levels of AUDIT-C enjoyed the game less and, via the indirect pathway, had lower levels of intention to drink less and drinking refusal self-efficacy. That achieving positive health outcomes is impeded in this manner for our sample's most hazardous drinkers (ie, those with high AUDIT-C levels) may relate to message discrepancy [51], which involves how the adoption of a recommended behavior (eg, decreased alcohol consumption) is most likely to occur when one's own current behavior is not that dissimilar from the recommended behavior. Thus, because the drinking behaviors of participants with higher scores on AUDIT-C are in sharp contrast with the game's recommended behavior, these participants are less likely to develop beneficial outcomes.

Limitations

We consider 4 main limitations here. First, while we assessed game time and risky alcohol decisions, there may be other in-game factors at play, including presence, flow, and engagement, which should be considered in future research. Second, our conceptual model included only one individual characteristic, AUDIT-C. Future research may want to use the full AUDIT because it has greater specificity in identifying hazardous drinking. Another individual factor that should be considered in future research is impulsivity, which is a well-recognized risk factor for both the initiation and continuation of alcohol misuse [52]. Third, another limitation of our study is that, while the one-group, pretest-posttest, quasi-experimental design permitted us to implement auto-regression models, an experimental design with a control group would be more robust. Fourth, we only assessed the short-term effects of the video game, but research suggests that the behavioral outcomes of serious health games may diminish with time [19]. To address this concern, it is recommended that future research measure both short-term and long-term outcomes.

Conclusions

Using the serious video game One Shot, we built and tested a multistep model for serious video game processes. The confirmation of research hypotheses suggests that serious video games can influence the alcohol consumption outcomes of young adults aged 21 to 25 years, which is an understudied at-risk group. Enjoyment is at the core of the documented processes, serving as an essential step in explaining how alcohol-related outcomes result from individuals' differential hazardous drinking levels and in-game decisions. This integral role of enjoyment has theoretical implications for researchers of serious video games and practical implications for game designers and developers.

Conflicts of Interest

None declared.

References

1. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Sep 22;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]

2. Patra J, Taylor B, Irving H, Roerecke M, Baliunas D, Mohapatra S, et al. Alcohol consumption and the risk of morbidity and mortality for different stroke types--a systematic review and meta-analysis. BMC Public Health 2010 May 18;10:258 [FREE Full text] [doi: 10.1186/1471-2458-10-258] [Medline: 20482788]

3. Rehm J, Shield KD, Roerecke M, Gmel G. Modelling the impact of alcohol consumption on cardiovascular disease mortality for comparative risk assessments: an overview. BMC Public Health 2016 Apr 28;16:363 [FREE Full text] [doi: 10.1186/s12889-016-3026-9] [Medline: 27121289]

4. Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. Prev Med 2004 May;38(5):613-619. [doi: 10.1016/j.ypmed.2003.11.027] [Medline: 15066364]

5. NIAAA Council Approves Definition of Binge Drinking. NIAAA Newsletter. 2004. URL: https://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.pdf [accessed 2020-11-13]

6.   Esser MB, Hedden SL, Kanny D, Brewer RD, Gfroerer JC, Naimi TS. Prevalence of alcohol dependence among US adult drinkers, 2009-2011. Prev Chronic Dis 2014 Nov 20;11:E206 [FREE Full text] [doi: 10.5888/pcd11.140329] [Medline: 25412029]

7.   Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]

8.   Rodriguez DM, Teesson M, Newton NC. A systematic review of computerised serious educational games about alcohol and other drugs for adolescents. Drug Alcohol Rev 2014 Mar;33(2):129-135. [doi: 10.1111/dar.12102] [Medline: 24329810]

9.   Thompson D, Baranowski T, Buday R, Baranowski J, Thompson V, Jago R, et al. Serious Video Games for Health How Behavioral Science Guided the Development of a Serious Video Game. Simul Gaming 2010 Aug 01;41(4):587-606 [FREE Full text] [doi: 10.1177/1046878108328087] [Medline: 20711522]

10.   Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]

11.   Boendermaker WJ, Veltkamp RC, Peeters M. Training Behavioral Control in Adolescents Using a Serious Game. Games Health J 2017 Dec;6(6):351-357. [doi: 10.1089/g4h.2017.0071] [Medline: 28926286]

12.   Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a Web-Based Computer-Tailored Game to Reduce Binge Drinking Among Dutch Adolescents: A Cluster Randomized Controlled Trial. J Med Internet Res 2016 Feb 03;18(2):e29 [FREE Full text] [doi: 10.2196/jmir.4708] [Medline: 26842694]

13.   Shandley K, Austin D, Klein B, Kyrios M. An evaluation of 'Reach Out Central': an online gaming program for supporting the mental health of young people. Health Educ Res 2010 Aug;25(4):563-574 [FREE Full text] [doi: 10.1093/her/cyq002] [Medline: 20150530]

14.   Willmott T, Russell-Bennett R, Drennan J, Rundle-Thiele S. The Impact of Serious Educational Gameplay on Adolescent Binge Drinking Intentions: A Theoretically Grounded Empirical Examination. Health Educ Behav 2019 Feb;46(1):114-125. [doi: 10.1177/1090198118780493] [Medline: 30027760]

15.   Patrick ME, Terry-McElrath YM, Lanza ST, Jager J, Schulenberg JE, O'Malley PM. Shifting Age of Peak Binge Drinking Prevalence: Historical Changes in Normative Trajectories Among Young Adults Aged 18 to 30. Alcohol Clin Exp Res 2019 Feb;43(2):287-298 [FREE Full text] [doi: 10.1111/acer.13933] [Medline: 30645773]

16.   Britton A, Ben-Shlomo Y, Benzeval M, Kuh D, Bell S. Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. BMC Med 2015 Mar 06;13:47 [FREE Full text] [doi: 10.1186/s12916-015-0273-z] [Medline: 25858476]

17.   Duggan M. Gaming and gamers. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/12/15/gaming-and-gamers/ [accessed 2020-11-13]

18.   Harteveld C, Bekebrede G. Learning in Single-Versus Multiplayer Games: The More the Merrier? Simulation & Gaming 2010 Aug 05;42(1):43-63. [doi: 10.1177/1046878110378706]

19.   DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]

20.   Klimmt C, Hefner D, Vorderer P. The Video Game Experience as "True" Identification: A Theory of Enjoyable Alterations of Players' Self-Perception. Communication Theory 2009;19(4):351-373. [doi: 10.1111/j.1468-2885.2009.01347.x]

21.   Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change1. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]

22.   Hong T, Beaudoin CE, Johnson C. A panel study of peer norms and adolescent alcohol consumption: developing strategies for communication interventions. J Health Commun 2013 Aug;18(8):913-930. [doi: 10.1080/10810730.2013.768729] [Medline: 23767700]

23.   Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA. Social Drinking on Social Media: Content Analysis of the Social Aspects of Alcohol-Related Posts on Facebook and Instagram. J Med Internet Res 2018 Jun 22;20(6):e226 [FREE Full text] [doi: 10.2196/jmir.9355] [Medline: 29934290]

24.   Tolchinsky A, Jefferson SD. Problematic video game play in a college sample and its relationship to time management skills and attention-deficit/hyperactivity disorder symptomology. Cyberpsychol Behav Soc Netw 2011 Sep;14(9):489-496. [doi: 10.1089/cyber.2010.0315] [Medline: 21288135]

25.   Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]

26.   Williams D. The Mapping Principle, and a Research Framework for Virtual Worlds. Communication Theory 2010;20(4):451-470. [doi: 10.1111/j.1468-2885.2010.01371.x]

27.   Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 2007 Jul;31(7):1208-1217. [doi: 10.1111/j.1530-0277.2007.00403.x] [Medline: 17451397]

28. Camchong J, Endres M, Fein G. Decision making, risky behavior, and alcoholism. Handb Clin Neurol 2014;125:227-236. [doi: 10.1016/B978-0-444-62619-6.00014-8] [Medline: 25307578]

29. Vorderer P, Klimmt C, Ritterfeld U. Enjoyment: At the Heart of Media Entertainment. Commun Theory 2004 Nov;14(4):388-408. [doi: 10.1111/j.1468-2885.2004.tb00321.x]

30. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.

31. Przybylski AK, Ryan RM, Rigby CS. The motivating role of violence in video games. Pers Soc Psychol Bull 2009 Feb;35(2):243-259. [doi: 10.1177/0146167208327216] [Medline: 19141627]

32. Boyle EA, Connolly TM, Hainey T, Boyle JM. Engagement in digital entertainment games: A systematic review. Computers in Human Behavior 2012 May;28(3):771-780. [doi: 10.1016/j.chb.2011.11.020]

33. Ruiter RA, Abraham C, Kok G. Scary warnings and rational precautions: A review of the psychology of fear appeals. Psychology & Health 2001 Nov;16(6):613-630. [doi: 10.1080/08870440108405863]

34. Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.

35. Oei TP, Hasking PA, Young RM. Drinking refusal self-efficacy questionnaire-revised (DRSEQ-R): a new factor structure with confirmatory factor analysis. Drug Alcohol Depend 2005 Jun 01;78(3):297-307. [doi: 10.1016/j.drugalcdep.2004.11.010] [Medline: 15893161]

36. Foster DW, Dukes K, Sartor CE. The road to drink is paved with high intentions: Expectancies, refusal self-efficacy, and intentions among heavy drinking college students. Alcohol 2016 Feb;50:65-71 [FREE Full text] [doi: 10.1016/j.alcohol.2015.11.004] [Medline: 26810041]

37. Ho SS, Lwin MO, Sng JR, Yee AZ. Escaping through exergames: Presence, enjoyment, and mood experience in predicting children's attitude toward exergames. Computers in Human Behavior 2017 Jul;72:381-389. [doi: 10.1016/j.chb.2017.03.001]

38. Cohen EL. Enjoyment of a counter-hedonic serious digital game: Determinants and effects on learning and self-efficacy. Psychology of Popular Media Culture 2016 Apr;5(2):157-170. [doi: 10.1037/ppm0000052]

39. Lyons EJ, Tate DF, Komoski SE, Carr PM, Ward DS. Novel approaches to obesity prevention: effects of game enjoyment and game type on energy expenditure in active video games. J Diabetes Sci Technol 2012 Jul 01;6(4):839-848 [FREE Full text] [doi: 10.1177/193229681200600415] [Medline: 22920810]

40. Cook TD, Campbell DT. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston, MA: Houghton Mifflin; 1979.

41. Alonso Puig M, Alonso-Prieto M, Miró J, Torres-Luna R, Plaza López de Sabando D, Reinoso-Barbero F. The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology. J Med Internet Res 2020 Mar 30;22(3):e16013 [FREE Full text] [doi: 10.2196/16013] [Medline: 32224482]

42. Aalbers T, Qin L, Baars MA, de Lange A, Kessels RP, Olde Rikkert MG. Changing Behavioral Lifestyle Risk Factors Related to Cognitive Decline in Later Life Using a Self-Motivated eHealth Intervention in Dutch Adults. J Med Internet Res 2016 Jun 17;18(6):e171 [FREE Full text] [doi: 10.2196/jmir.5269] [Medline: 27317506]

43. Phan MH, Keebler JR, Chaparro BS. The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). Hum Factors 2016 Dec;58(8):1217-1247. [doi: 10.1177/0018720816669646] [Medline: 27647156]

44. Planken M, Boer H. Effects of a 10-minutes peer education protocol to reduce binge drinking among adolescents during holidays. Journal of Alcohol and Drug Education 2010;54(2):35-52.

45. Belgrave FZ, Reed MC, Plybon LE, Corneille M. The impact of a culturally enhanced drug prevention program on drug and alcohol refusal efficacy among urban African American girls. J Drug Educ 2004;34(3):267-279. [doi: 10.2190/H40Y-D098-GCFA-EL74] [Medline: 15648887]

46. Ullman JB. Structural equation modeling. In: Tabachnik LS, Fidell BG, editors. Using Multivariate Statistics. New York, NY: Harper Collins; 1996:709-811.

47. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]

48. Finkel SE. Causal Analysis with Panel Data. Thousand Oaks, CA: Sage; 1996.

49. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002 Mar;7(1):83-104 [FREE Full text] [doi: 10.1037/1082-989x.7.1.83] [Medline: 11928892]

50. Craig CC. On the Frequency Function of xy. Ann. Math. Statist 1936 Mar;7(1):1-15. [doi: 10.1214/aoms/1177732541]

51. 2008 World Population Data Sheet. Population Reference Bureau. 2008. URL: https://www.prb.org/2008wpds/ [accessed 2020-11-13]

52. Dick DM, Smith G, Olausson P, Mitchell SH, Leeman RF, O'Malley SS, et al. Understanding the construct of impulsivity and its relationship to alcohol use disorders. Addict Biol 2010 Apr;15(2):217-226 [FREE Full text] [doi: 10.1111/j.1369-1600.2009.00190.x] [Medline: 20148781]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Concise
CFA: confirmatory factor analysis
CFI: comparative fit index
PMT: protection motivation theory
RMSEA: root mean square error of approximation
RQ: research question
SEM: structural equation model
SSI: Survey Sampling International

XSL·FO
RenderX

<u>Original Paper</u>

Using Online Support Communities for Tourette Syndrome and Tic Disorders: Online Survey of Users' Experiences

Victoria Perkins[1], BSc, MSc; Neil S Coulson[1], PhD; E Bethan Davies[2,3], BSc, MSc, PhD

[1]Division of Rehabilitation, Ageing and Wellbeing, School of Medicine, University of Nottingham Medical School, Queen's Medical Centre, Nottingham, United Kingdom

[2]Division of Psychiatry & Applied Psychology, Institute of Mental Health, The University of Nottingham, Nottingham, United Kingdom

[3]NIHR MindTech MedTech Co-operative, Institute of Mental Health, The University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
E Bethan Davies, BSc, MSc, PhD
NIHR MindTech MedTech Co-operative, Institute of Mental Health
The University of Nottingham
Triumph Road
Nottingham
United Kingdom
Phone: 44 0115 748 4098
Email: bethan.davies@nottingham.ac.uk

Abstract

Background: People living with a tic disorder (TD)—such as Tourette syndrome (TS)—experience many negative psychological and social challenges arising from chronic tics, such as stigmatization from peers and poorer quality of life, and these can impact upon their families too. It can be difficult for this population to access face-to-face support for tics, and so online support communities offer one avenue for support from peers facing similar experiences. However, little is known about how online support communities may be used by people with TS and other TDs, and by others (eg, parents, caregivers) supporting a person with TS/TD.

Objective: This study aimed to explore users' experiences of participation in online support communities for TS and TDs.

Methods: In total, 90 respondents (aged 13-62 years; 62% [56/90] female) from 13 countries completed an online survey exploring their experiences of using online support communities for TS and TDs. Respondents were people living with TS/TD themselves (n=68) or supportive others of someone with TS/TD (eg, parent, sibling, spouse; n=14), or both (n=8). The online survey contained open-ended questions eliciting their self-reported motivations for using online communities, their benefits and drawbacks of participation, and whether online support communities affected offline management of tics. Responses were analyzed using thematic analysis.

Results: Seven overarching themes captured experiences of using online support communities for TS/TDs. The overwhelming reason for their use was to find accessible support due to a lack of offline face-to-face support. Online support communities were valued sources of informational and emotional support, and also had a positive impact upon helping users' psychological well-being. Online communities helped provide a space where people with TS/TDs could feel accepted and reduce the social isolation they felt offline. The suggestible nature of tics and being reminded of the challenging nature of TDs were main disadvantages arising from using online support communities, alongside conflict arising within online communities.

Conclusions: The findings suggest that online support communities appear to offer valuable informational and emotional support to those living with TS/TD and their families too, especially given the lack of locally available support. This facilitates a sense of community online, which can help users in overcoming long-standing social isolation and aid self-reported improvements in psychosocial well-being. Users reported some drawbacks in engaging with online support communities, such as conflict between different types of users and triggering content, which negatively affected experiences of community participation.

XSL•FO

RenderX

Introduction

Tic disorders (TDs)—such as Tourette syndrome (TS)—are noncurable neurodevelopmental conditions characterized by sudden, persistent, purposeless motor movements, or vocalizations known as tics [1]. TDs typically have their onset during childhood, and tend to peak in early adolescence [2]. Transient TD (ie, tics present for <1 year) is the most common, affecting approximately 3% of children; chronic TD (motor or vocal tics present for >1 year) and TS (motor and vocal tics for >1 year) affect approximately 1.61% and 0.77%, respectively [3]. Tics fluctuate in both frequency and severity over time, and for many tend to decline with age [4].

The processes involved in the assessment, diagnosis, and treatment of TDs can be tricky to navigate: delays in referral and diagnosis are common due to the fluctuating nature of tics, and how they can be indicative of other health issues (eg, eye problems, allergies) [5]. Patients with TDs describe many health care professionals as being uninformed about tics, in turn contributing to delays in diagnosis and intervention [6]. Furthermore, environmental and systemic barriers—such as geographical location, issues with funding, and service delays—between specialist services and patients often hinder access [7,8]. Evidence-based pharmacological and behavioral treatments and supportive care strategies can facilitate reductions in tics [9,10]. Pharmacological and behavioral treatments are effective for TDs but with unique issues affecting their use: for example, medications are often accompanied by adverse effects (eg, fatigue, drowsiness, weight gain) [7,11]. Additionally, TDs are seldom seen in isolation; attention deficit hyperactivity disorder, obsessive-compulsive disorder, autism spectrum disorder, and learning disorders are common comorbidities [12,13]. These comorbidities can further affect assessment and treatment options, as well as further impacting upon quality of life (QoL).

Chronic tics can impair functioning across many areas of everyday life for children and adults [2]. This includes their physical impact (eg, pain and injury from tics), emotional and psychological impact (eg, increased anxiety and hopelessness, questioning sense of self), impact upon school and work (eg, tics impacting upon classroom learning, misunderstanding from teachers, finding employment), and impact upon social life and relationships (eg, bullying, affecting family dynamics) [2,14,15]. TDs also impact upon the family too: interviews with parents of children with TS describe having to cope daily with tic-related challenges, such as seeing their child have painful tics, disruption to daily routines, and financial difficulties arising from additional caring responsibilities and tic-related behaviors [16]. The visible nature of tics often attracts attention from other people [17], and the negative impact of chronic tics on social relationships is widely known: stigmatization from others, discrimination, isolation, bullying, and social exclusion arising from TDs are common interpersonal issues found across the life span [6,18]. The consequences of childhood rejection and stigma can be long-lasting, with adults in one study describing purposely avoiding others to evade ridicule [18].

Social support is an important resource in coping with chronic conditions [19], and may help lessen the detrimental impact of TDs upon psychosocial well-being of individuals with TD and their families [20]. Receiving social support from others with similar lived experiences can help reduce isolation and aide adjustment and coping with tics [16,18,21]. Wadman et al [17] report how young people may find it difficult talking to their non-TS peers about tics—for example, being concerned about others' reactions and perceptions—and suggest learning about other strategies people with TS use to educate others could be beneficial. Social support is important for families too: in an Australian study involving mothers of children with TS, respondents described the social isolation and decline in social support arising from the challenges of TS, and placed particular importance on the value of making connections with people who understood the lived experience of TS—such as other parents [21]. Furthermore, increased perceived social support among parents of children with TD was associated with emotional and practical coping mechanisms that helped manage tic-related difficulties [20]. However the relative rarity of TDs means local support—particularly for adults with TD—is seldom accessible [18].

With increasing access to the internet, online support communities provide new opportunities for people affected by specific health issues to share, communicate, and network together in an online space [22,23]. Online support offers many features that are not available in face-to-face offline support, including reducing geographical and financial restrictions in access, greater anonymity in participation, and being able to access informational and emotional support that may not be available offline [22,24]. The content of online support communities is generated by the users themselves, meaning they can discuss issues important to them. Online support communities are typically asynchronous platforms (eg, online forums on charity websites) and in recent years disease-specific groups have emerged on social networking sites (eg, Facebook) [23,25]. Furthermore, participation in online support communities can be more *active* or *passive*: users may actively engage in the community (eg, post messages, respond to others, upvote/like others' content) or be more passive in reading others' content but not interacting with it [26]. Benefits and drawbacks of online support community participation may vary by the type of chronic illness support is sought for; for example, women using a polycystic ovary syndrome online support community reported feeling empowered by reading about others living fulfilled lives despite their condition [27], while users of an inflammatory bowel disease online support community found peer support useful in guiding their treatment decisions [28]. For young people, online support communities can help validate and normalize what they are experiencing, sharing their thoughts, feelings, and experiences with similar others [24]. Online support communities are not without their problems and a range of difficulties have been described in the literature: greater anonymity and distance mean users may be rude to others, the wealth of information can be difficult to process and inaccurate information can be shared without fact checking, and the absence of facial and vocal cues in online communication can lead to misunderstandings in support [29]. However,

whether there exists any problems arising from engagement with TD online support communities remains to be seen.

For individuals with TDs and their support networks, online support communities could potentially provide a viable outlet to find social support from peers facing similar tic-related challenges, which subsequently may facilitate management of tics and aid their psychosocial well-being. People with TS have described how they may conceal and suppress their tics in public [17,18,30]: people with TS may have more control over how they present themselves online, and tics (and common comorbidities) may interfere less in seeking out online support compared with face-to-face. Previous research has highlighted the importance of peer acceptance and supportive friendships in providing practical and emotional support [6], but to date there has been little exploration of social support from peers with TDs. Therefore the aim of this study is to explore users' perceptions and experiences of using online support communities for TDs, in order to gain insight into their participation in online support communities, benefits and drawbacks of being using online support communities, and the impact of using online support communities upon their offline management of TDs. Users of these online support communities reflect those with a TD (diagnosed or suspected), and supportive others (eg, parents, siblings, spouses and friends).

Methods

Participants and Recruitment

The study involved recruiting users of online support communities for TDs and TS. Relevant online support communities were identified via searches using Google and on Facebook employing a range of key terms in different combinations, such as *online support*, *online group* and *tics*, *tic disorders*, or *Tourette syndrome*. As there had been previous little research exploring online support communities for TDs/TS, it was felt that the criteria outlined by Coulson et al [23] provided a methodical way to identify suitable and active online support communities for this study. To identify eligible online support communities, they had to be (1) user led (ie, not led by health professionals) and *currently active*, meaning they had a minimum of 30 postings per month; (2) in English language; and (3) the terms and conditions of the online support communities did not preclude research. In online support communities where administrators/moderators were identifiable and contactable, an introductory message was sent to them requesting permission to post study information in the online support community. Where no administrators/moderators were apparent, the lead researcher (VP) independently posted the study advert directly to the online support community. Together, 14 eligible online support communities were contacted for recruitment purposes (n=11 Facebook groups, n=3 forums on other websites). Permission to recruit was obtained from 4 online support communities, with the remaining either declining to participate (n=7; reasons for declining were not provided) or failing to respond (n=3). Additionally, the advert was circulated by Tourettes Action (a UK-based TS charity) via their newsletter and social media channels, and was shared via Twitter by the NIHR MindTech MedTech Co-operative.

After identifying eligible online support communities, to take part in the study participants were (1) people with TD/TS themselves (diagnosed or suspected) or be someone supporting a person with TD/TS (eg, a parent, sibling, partner, or friend), or could be in both categories; (2) aged 13 years or older, as is typically the minimum age legally required to register to use online support communities [28]; and (3) be a member of 1 (or more) online support community specifically for TDs.

The study adverts provided information about the study aims and invited individuals to click on a link to the online survey (hosted on Jisc Online Surveys). On the survey's landing page, respondents were presented with a more comprehensive overview of the study and their legal rights and protections. Participants were required to confirm they had read and understood the information and indicated their willingness to participate through completing an online consent form. Next, participants completed questions about themselves, their lived experience of TD/TS, and self-reported use of online support communities, as well as 6 open-ended questions exploring their personal motivations for using online support communities, benefits and drawbacks of participation, and ways in which the online support had impacted upon offline management of TDs (Textbox 1). The questions were based on those from previous research exploring online support community use for different health issues and chronic conditions [23,27,28,31,32]. Each question was accompanied by a free-text box, providing participants the opportunity to share their experiences.

Textbox 1. Open-ended questions presented in the online survey.

1. What first led you to use an online community for Tourette syndrome/tics?

2. What, if any, do you consider to be the main benefits to participating in online communities for Tourette syndrome and/or tics?

3. What, if any, drawbacks have you encountered in participating in online communities for Tourette syndrome and/or tics?

4. How have online communities influenced the way you manage your tics, or how you help the person you support (child/sibling/spouse/friend) with their tics? If so, please give some examples.

5. Has anything you have been made aware of or learnt from online communities helped you or the person you support (child/sibling/spouse/friend) in any other way? If so, please give some examples.

6. Is there anything else you wish to tell us about your experience of using online communities for Tourette syndrome and/or tics?

Patient and Public Involvement

Prior to finalizing the 6 open-ended questions, potential questions were discussed with a member of the Patient and Public Involvement (PPI) group from the Online Remote Behavioural Interventions for Tics trial (ORBIT trial [33]): the PPI group members have a diagnosis of TS or are a parent to a child with TS. Topics arising from this conversation were used to structure the final set of questions. The full online survey was subsequently reviewed by 3 members of the PPI panel to check it was understandable, appropriate, and sensitive to its intended audience.

Ethical Considerations

The study was reviewed and approved by the University of Nottingham Division of Rehabilitation, Ageing & Wellbeing Ethics Committee. The first 2 pages of the online survey consisted of comprehensive study information, rights in participation, and withdrawal procedures, alongside the research team's contact details. The online consent form consisted of several statements, all of which the participant needed to agree to in order to participate in the study. All responses were kept securely: access to data was granted to members of the research team only, and all data were downloaded to and stored on a password-protected computer. No participants chose to withdraw after completing the online survey.

Analysis

Responses to the open-ended questions were analyzed by the first author (VP), using an inductive approach to thematic analysis [34]. As the role of online support communities for TDs had not been explored previously in research, a data-driven inductive approach without a theoretically derived coding scheme was judged as a practical analytic strategy to explore all aspects of the data. Analysis followed Braun and Clarke's [34] guidelines: first, data familiarization and searching for initial meanings were performed through reading and re-reading each survey response several times. Second, preliminary codes detailing basic interpretation of responses were made, and similarly coded data were grouped together. Next, similarly coded data were organized into related clusters, which generated initial theme names. These initial themes were reviewed and the data used to support them examined. In instances where evidence gathered was insufficient to support a theme, codes were regrouped. Themes were defined and refined into subthemes to ensure findings credibly captured all important elements of the data set. o support this process, the second (NC) and third (EBD) authors reviewed the themes/subthemes and coded data, and with the first author (VP) finalized the themes.

Results

Participant Demographics

A total of 90 participants completed the online survey (Table 1): the majority (68/90, 76%) had a TD themselves. Slightly less than two-thirds of the sample were female (56/90, 62%), with ages ranging from 13 to 62 years (mean 27.28 [SD 13.11] years), and the majority resided in the United States (38/90). Participants varied in how long they (or the person they supported) had been living with tics, ranging from 6 months to 50 years (mean 12.56 [SD 13.19] years; median 9 years). Self-reported comorbidities were common, with over four-fifths (74/90, 82%) reporting at least one comorbid condition. The majority were members of more than 1 online support community (68/90, 76%) and reported using them for less than a year (53/90, 59%). Many respondents self-reported accessing online support communities frequently, with one-third accessing online support communities everyday (32/90) and another third accessing several times a week (34/90).

Thematic analysis generated 7 overarching themes reflecting reasons for accessing online support, the informational and emotional support gained from online support communities, feeling accepted and reducing isolation in having TDs, and problems with using online support.

Table 1. Demographic composition of sample (N=90).

Demographic	Value
Gender, n (%)	
Male	25 (28)
Female	56 (62)
Nonbinary	6 (7)
Other	2 (2)
Age, mean (SD)	
Person with TD[a] (n=68)	24.04 (11.32)
Person with TD who also supports someone with TD (n=8)	31.38 (16.07)
Non-TD person who supports someone with TD (n=14)	40.64 (10.93)
Location, n (%)	
Australia	8 (9)
Austria	1 (1)
Belgium	3 (3)
Canada	5 (6)
Germany	5 (6)
Ireland (Republic of)	1 (1)
Italy	2 (2)
Netherlands	1 (1)
New Zealand	1 (1)
Norway	1 (1)
Poland	1 (1)
United Kingdom	23 (26)
United States	38 (42)
Self-reported co-occurring conditions[b], n (%)	
Autism spectrum disorder	22 (24)
Attention deficit hyperactivity disorder	27 (30.0)
Obsessive-compulsive disorder	41 (46)
Anxiety	62 (69)
Depression	39 (43)
Dyslexia or other learning difficulty	13 (14)
Dyspraxia	6 (7)
Other	13 (14)
Membership of TS[c]/TD online support communities, n (%)	
Member of one online support community	21 (23)
Member of two or more online support communities	69 (77)
Frequency of using TS/TD online support communities, n (%)	
Every day	32 (36)
Several times a week	34 (38)
Once a week	10 (11)
Once every couple of weeks	5 (6)
Once a month	3 (3)

Demographic	Value
Once every couple of months	1 (1)
Once a year	2 (2)
Length of using online support communities for TS/TD, n (%)	
0-6 months	33 (37)
7-12 months	20 (22)
1-2 years	12 (13)
More than 2 years	25 (28)

[a]TD: tic disorder.

[b]Values add up to more than 100% as participants could self-report more than one co-occurring condition.

[c]TS: Tourette syndrome.

Using the Internet to Find Accessible Support

This theme describes how minimal availability of face-to-face support encouraged participants to seek out support from peers online.

Lack of Offline Support From Peers and Health Care Practitioners

Participants discussed how lack of in-person support—from people with TD and from health professionals—for TDs motivated their use of online platforms to find support.

> *I want to interact with others who also have the condition (…) adults have so few options for support.* [P40, aged 23, has TD]

Participants described their wish to access support from others experiencing TDs, and their dissatisfaction regarding the limited face-to-face opportunities to do so. Additionally implied is that many health care professionals appeared to have a basic understanding of TDs, but lacked a thorough appreciation of the difficulties faced by their patients. Finding and using online support communities was seen as a viable option to aid this.

> *Medical professionals need to realise that people go to these forums more than their doctors because your standard GP etc has no idea what TS is really like, we provide our own support because no one else will. We are like a forgotten element of the medical society.* [P27, aged 43, has TD]

Breaking Geographical Barriers

Despite some discontent regarding limited face-to-face support, multiple participants appreciated how online groups meant they could access support regardless of their geographical location to account for their limited face-to-face support from others with TDs:

> *Such a valuable resource when don't have any local TS families.* [P24, aged 46, parent of child with TD]

While face-to-face support groups may provide small local networks for patients, online support communities facilitate cross-cultural interactions providing people greater choice and opportunity to find and communicate with similar peers.

Ease of Access

Being able to access support via online support communities at any time was considered beneficial compared with scheduled face-to-face groups, especially by those who felt that accessing face-to-face support was challenging:

> *I actually prefer the online connections as opposed to in person because I have anxiety and the internet makes it easier for me to talk to people.* [P44, aged 19, has TD]

A small number of participants also advocated for online communication as helpful when comorbid conditions hindered their attendance at face-to-face groups.

Online Communities as Informational Sources

This theme explores how information obtained in online support communities supports users' offline management of TDs and common comorbidities. Experienced online support community users were depicted as *experts* by participants, well-known for providing practical advice and strategies for lessening the impact of TDs.

Signposting to Health Services and Diagnostic Advice

Signposting newer members to appropriate health services and charities was commonly reported by participants, and had assisted many in acquiring a TD diagnosis and guidance on ways to explain tics and comorbidities to others. For many participants, the decision to access online support communities was motivated by concerns they may have a TD:

> *My initial suspicions that I might have TS led me to search for online communities (…) that community convinced me to speak to a doctor.* [P11, aged 50, has TD]

Helpful in Managing Tics

Several participants described how they tried out strategies shared by online support community members, such as learning to *redirect* tics, to manage their tics:

> *I've learned how to shorten my tic attacks, tricks and tips for painful tics like a head pillow for neck jerking and padded gloves for self-hitting tics.* [P13, aged 20, has TD]

One use of online support communities by participants was to share acquired behavioral and lifestyle management techniques with struggling, newly diagnosed members. Participants described how online support communities can facilitate TD management by allowing users to seek out personalized advice pertaining to specific, bothersome, and often self-injurious tics.

Understanding Common Comorbidities to Gain Personal Insight

Information shared regarding common comorbidities appeared to reassure and comfort users (both people with TD and parents supporting a child) about the commonness of comorbidities with and the complexities involved in managing multiple conditions:

> The awareness of the comorbidities that come alongside Tourettes has helped me and my family be less confused and worried as to why I had such bad OCD. [P25, aged 16, has TD and sibling with TD]

Explaining Tics to Other People

Respondents described how reading other users' accounts of similar discussions aided how they communicated their TD to others, and used information obtained from online support communities to facilitate offline discussion with other people:

> I learned a lot of information about Tourette's, a lot of personal accounts that aligned with my experiences and comforted my worries, and ways to handle opening up about my TS to friends and family. [P80, aged 21, has TD]

Being signposted to informational resources designed for teachers and classmates was reported to have great benefit in helping schools provide effective support:

> It has helped me to find information for school on how to deal with tics and about how Tourette's symptoms manifest themselves. [P31, aged 53, parent of child with TD]

Patients Are Experts by Experience

Participants were keen to highlight the importance of acquiring information from people with lived experience of TD:

> Support & advice from people & other families who have the experience of living with TS. [P24, aged 46, parent of child with TD]

On occasion peer expertise was discussed by participants somewhat evocatively, denoted using capital letters or emotive language to emphasize the conviction underpinning the sentiment. This seemingly reflected a perceived need to accentuate the invaluable nature of patient voice in developing a nuanced understanding of TDs.

Online Communities as Emotional Support Sources

Respondents appreciated online support communities for providing an outlet for sharing and discussion between those experiencing similar difficulties. Mutuality was considered fundamental for other members to gain a comprehensive understanding of personal TD experiences. Such insight fostered emotional support and cultivated online environments in which

participants felt comfortable in speaking openly and able to build friendships with likeminded users.

Seeking Understanding Through Shared Experiences

Narratives of lived experiences were felt to convey more valuable insights into aspects of TDs that informational sources could not offer. People with TDs were considered to possess a much-desired level of empathy and relatedness that those in participants' offline world (eg, parents, siblings, non-TD friends) could not provide:

> Contact with others who know what it's like to live with TS, WE are the real experts. [P27, aged 43, has TD]

Sharing in a Safe and Supportive Space

People with TD and those supporting them reported that online support communities provided them the opportunity to express feelings freely and safely:

> It can help you get what you're feeling out. If you're stressed about something or need to rant you can. [P19, aged 16, has TD]

Inferred was that online support communities provide outlets for members to discuss ongoing difficulties unreservedly and without the *censorship* they may experience when discussing emotive events with people in their offline world. The internet-mediated nature of communication between online support community members seemingly allowed participants to speak honestly and receive unquestioning empathy and support without fear of ramification.

Forging Friendships and Bonding With Peers

Through receiving emotional support from likeminded others and communicating with people who understood the lived experience of TD, participants detailed how online support communities helped support the development of new friendships:

> These communities are a great place to find (...) friends who you can talk and play games with and support that extends beyond tics into general life problems. [P13, aged 20, has TD]

Feeling Part of a Community

Participation in supportive online support communities, particularly smaller communities, reportedly evoked a sense of community among members. The bonds formed with peers in online support communities were considered to replicate in-person support networks and facilitated connectedness and belonging to online environments.

Reducing a Sense of Social Isolation

People with TD and those supporting them both felt online support communities helped them feel less removed from society. Interacting with online support community members helped reduce feelings of loneliness, and many reported feeling less *abnormal* or *weird* as a result of using online support communities:

> Until I found the community, I was more or less convinced that I was alone in the way I felt – often

rejected and embarrassed about my tics. I don't feel like that anymore. [P10, aged 21, has TD]

For some, contact with individuals experiencing TDs was sufficient to feel integrated into a community, while others were inducted by reading members' TD experiences and drawing parallels to personal life events.

Acceptance and Belonging

Participants discussed being immersed in accommodating, nonjudgmental online environments, with closeness and camaraderie between group members facilitating a sense of belonging:

> *A wonderful [OSC] for Tourette's has helped so many people with accepting not only themselves but others and has made everyone a close family.* [P34, aged 17, has TD]

Connections between online support community members were depicted as akin to familial relationships, further illustrating the community atmosphere cultivated within online support communities.

Paying It Forward

Several participants reported that sharing their experiences and expertise were useful to newer, struggling online support community members. Sharing experiences to engage in a process of reciprocal support was commonly discussed:

> *If people are going through the same struggles I did when I was younger I want to offer advice.* [P84, aged 21, has TD]

Participants described a desire to *give back* to communities who had helped them throughout difficult times and a sense of validation of their abilities to support others. This seemingly perpetuated a cycle of reciprocal support that enabled better adjusted participants to remain connected to the group and in continued receipt of other psychosocial benefits.

Preference for Smaller Community

Some participants reported it being more difficult to find a sense of community and community benefits (friendship, closeness, and belonging) in larger online support communities:

> *It is hard to develop strong friendships or have conversations within groups of hundreds of members.* [P82, aged 17, has TD]

For many, the inability to engage with all online support community members was a drawback and some participants expressed concerns regarding the motivations of lesser known, often silent members. Overall, respondents typically felt more comfortable in and supported by smaller, intimate online support communities.

Positive Impact on Psychosocial Well-Being

This theme denotes the positive impact online support communities have on participants' attitudes and behaviors toward tics. Learning of others' TD experiences resulted in a shift in beliefs about tics; several respondents expressed gratitude that their tics were less functionally impairing than many users, while some benefitted from the normalizing impact

had by witnessing others live fulfilled lives despite tics. In turn this impacted participants' *offline* world, with many reporting increased social activity and improved self-confidence in the ability to withstand public responses to tics.

Reappraising Own Circumstances

Comparison of oneself with other group members was found beneficial to most participants; hearing about more impairing TD experiences seemingly aided participants in reappraising their perception of their own TD, often as being minor compared with others:

> *Some people struggle with more severe/less socially acceptable tics, such as coprolalia or complex motor tics, I feel like mine aren't such a big deal after all.* [P81, aged 20, has TD]

This awareness appeared to help respondents accept their tics as less bothersome or apparent to other people, and reportedly helped minimize self-consciousness about their tics.

Improved Confidence and Acceptance of Tics

Through encouraging respondents' acceptance of tics, online support communities were found to facilitate engagement in typical everyday activities (eg, work, going out):

> *I've learned that I shouldn't be so self-conscious when I tic, that if people ask just explain and move on.* [P35, aged 14, has TD]

Reducing shame and embarrassment about tics appeared to facilitate their acceptance of them. This revised outlook seemingly bolstered participants' confidence, enabling many to embrace opportunities to socialize and address their disorder directly in their offline world. Acceptance of tics further aided participants to develop their own personal narratives to explain their tics to others:

> *They have made me more confident to tic in public. For most of my life, I have suppressed tics in public, leading to the feeling that I am hiding a part of me that I am ashamed of, followed by tic attacks when I get home.* [P40, aged 23, has TD]

Normalization of TDs Provides Hope for the Future

Members with positive lived experiences were found to reassure participants that TDs do not necessarily impact upon long-term QoL:

> *Seeing other people with TS who are happy, successful and empowered helped me to embrace my TS as it showed me that my condition should give me strength and not hinder me.* [P22, aged 16, has TD]

Experienced and potentially well-adjusted online support community users were depicted as role models for younger members, capable of offering hope regarding the future by illustrating that patients with TD can achieve the same life goals as typically developing peers. Virtually meeting other people with TD living *normal* lives was depicted as inspiring and encouraged participants to grow and learn from challenging experiences to achieve positive futures.

XSL•FO

RenderX

Reduction in Self-Blame

Self-blame was denoted among a handful of predominantly younger online support community members, many who once viewed tics as a *punishment*. Online support communities were found to help overcome this perception: the mechanisms through which this occurred were not discussed, but other positive influences on psychosocial well-being (acceptance, normalization, and reappraisal of circumstances) occurring following exposure to other individuals with TD plausibly helped facilitate this reduced self-blame:

> *I don't feel that my tics are my fault. I am trying to convince myself that I'm not doing something wrong by having tics, and this community is helping.* [P10, aged 21, has TD]

Problems With Online Support

Both tic-related and internet-based problems with online support community access were described by participants. A frequently mentioned concern by respondents was the tendency for non-evidence–based advice to be discussed and potentially mislead vulnerable users. Another concern centred around posts of graphic tic-related injury, with participants describing how this appeared to lead to an uncontrollable influx of tics, which often compelled temporary withdrawal from online support communities. Furthermore, reading others' negative experiences perpetuated low mood and discouraged participants' optimism for the future.

Tics Are Suggestible

Online support communities were commonly described as *triggering*, meaning how reading about tics or watching videos of tics could lead to heightened tic severity, frequency, or development of new tics. Most considered this a minor disadvantage, advising that avoiding online support communities when tics are particularly bothersome is a suitable solution:

> *Sometimes you can pick up new tics, or reading about others will make your own tics increase in frequency. (...) if I am having a "bad tic day" I will not read the posts in the community.* [P14, aged 18, has TD]

Several participants suggested using *warnings* preceding the body of particularly descriptive or visual posts.

Continuous Reminders of a Lifetime of Challenges

Exposure to others' daily struggles reminded participants that managing TD was a necessary part of life. Negative experiences were discussed somewhat apathetically, with a handful of participants seemingly resigned to a potentially restricted future and accepting difficulties they/the person they support may someday encounter. For some this perpetuated low mood and enhanced the perceived magnitude of TD-related problems:

> *It can be upsetting to see and read about the negative experiences some have had and can be depressing to be part of a group when you really just wish your child didn't have this extra issue to cope with.* [P56, aged 52, parent of child with TD]

How Sound Is Others' Advice?

Several participants expressed concerns regarding the validity of advice offered in online support communities:

> *There are many people who come looking for advice- but a lot of the advice given is misleading, inaccurate and sometimes dangerous.* [P58, aged 35, parent of child with TD]

Respondents recognized that particularly among parents of children with TD, desperation to improve their child's QoL led to them being vulnerable to accepting poor advice from unverified sources. Alternative treatment regimens were considered with skepticism, prompting multiple participants to emphasize that online support communities cannot replace medical guidance and recommend others to avoid engaging with experimental or anecdotal therapies.

Within-Group Conflict

Perceived judgment and a lack of consideration among online support community members were found to foster conflict between users, culminating in a negative community atmosphere. Many participants discussed encounters with unfriendly individuals within online support communities, citing intolerance for personal lifestyle choices and TD management styles as common reasons for withdrawing from online support communities.

Unfriendliness of Some Community Members

Respondents felt annoyed by those who expressed entitlement to greater empathy than others; this resulted in somewhat competitive conversations, not conducive to the supportive atmosphere that most members endeavored to create. Participants felt equal respect and acceptance were not afforded to all users:

> *Some people can be judgemental. In some groups people seem to try to compete for 'whose Tourette's is the worst'.* [P39, aged 27, has TD]

Several participants additionally detailed how insensitive or provocative comments were easily misconstrued and prematurely judged some users. A small number of participants reported experiencing cyber bulling or harassment, contributing to their withdrawal from otherwise supportive networks.

Rigidity of Community Perspective

Many recognized that online support communities typically possess a unified and unwavering perception of different approaches for managing tics:

> *Certain communities tend to develop a common voice. This one strongly supports CBD, that one strongly rejects it. This one loves CBIT, that one is sceptical. (...) either way one isn't getting a healthy, integrated community or understanding of TS.* [P18, aged 38, parent of child with TD and spouse of person with TD]

For some, rigid online support community opinions were considered disadvantageous. Several respondents experienced rejection upon initiating discussions regarding topics which did not align with an online support community's outlook.

Additionally, participants commented that newly diagnosed and possibly inexperienced patients with TD may only participate in one online support community, meaning they may only be exposed to certain perspectives or a few resources.

Conflict Between Different Types of Users

Frustrations regarding the tendency for parents of younger patients with TD to monopolize groups were common and discussed somewhat evocatively. Conflict between users with a TD and users supporting a person with TD were reported:

> *I find that parents of children with TS tend to have the loudest voices in some communities. It can focus conversations away from individuals who actually have the condition.* [P40, aged 23, has TD]

Parents were depicted as insensitive and inconsiderate toward other group users; not only did people with TD feel parent users undervalued their opinions, but many also considered their domineering presence in groups prevented adults with TD from accessing comprehensive social support. Moreover, participants found parents' use of online support communities as outlets to discuss the negative impact of their child's TD on family life particularly distressing in questioning their own family's true feelings toward them.

Discussion

Principal Findings

This study explored experiences of online support community users for those living with TDs and people supporting them, such as parents. Seven overarching themes were identified, many of which align with previous research into online support communities for long-term health conditions (eg, [23,27]) and appear to map onto empowering processes and outcomes attained through online communities [35,36], including having increased knowledge about TD, helping shape communication with health professionals, and guiding offline management of tics. Research into social support in TDs has typically looked at support from non-TD peers [17,37], has highlighted the value of support in parents of children with TD [16,21], and investigated social adjustment difficulties in this group [38]. This study adds to the limited literature looking at how people with TD may use online support from peers with TD (or from caregivers of people with TD), and the impact of this upon their tics and well-being.

Our findings suggest several unique benefits of online support communities for TDs. Online groups granted much-desired access to other people with TD, and allowed for diverse cross-cultural connections too. The continuous availability of online support communities allows for support regardless of time boundaries and accessibility to users whose comorbidities deterred attendance to face-to-face support groups. Given the prevalence of morbidities among people with TD [13], this may be particularly useful in recommending online support communities to patients and their families for whom attending in-person groups is particularly psychologically demanding.

Supporting previous literature [6], many respondents felt that medical professionals had inadequate TD knowledge—and so

online support communities helped users in gaining personalized guidance and experiential information from fellow peers. This suggests that online support communities can supplement medical advice and offer users sufficient information to manage tics autonomously. Moreover, discussion regarding comorbidities was considered enlightening to participants and provided explanations for long-standing differences between themselves and non-TS peers. This normalization of experiences and sharing of resources appeared to help respondents build their confidence and encouraged initiation of TD-related conversations with others offline (eg, friends, family, teachers). Although online support communities perhaps emphasize the limitations of health care and public TD awareness, their provision of experiential information facilitates better offline management of tics and associated difficulties.

By contrast, 2 drawbacks of online support communities reported by some respondents were the credibility of advice provided by online support community users and rigidity of certain communities in terms of having strong opinions regarding different types of tic management. This could result in users finding it difficult to discuss evidence-based treatments that might not align with the online support community's preferences, or may be directed toward non-evidence–based treatments. While the role of a moderator/administrator is generally consistent across online support communities (eg, to establish and maintain ground rules through overseeing group communication to maintain safe online space), each community will vary in how they moderate their online space [39], including the discussion of evidence-based and alternative treatments. Voluntary moderators of online support communities are often patients themselves or caregivers of patients [39], and their personal experiences and beliefs about tic management could potentially impact upon how they moderate online support communities. Shoebotham and Coulson [40] suggested that having health professionals as moderators could help in appraising (mis)information posted in online support communities; however, online support communities may wish to be separate from external organizations, and may not want professionals in their patient-led online spaces.

Emotional support was perhaps the greatest incentive for continuing to engage in online support communities for TDs. Aligning with online support communities for other chronic conditions [27,28], respondents benefitted from encountering likeminded others whose mutuality seemingly facilitated a level of understanding not apparent among non-TD networks. This was characterized by distinctions between sympathy and empathy, and provided a foundation for connections otherwise lacking in encounters with non-TD peers. Comfort in being surrounded by people who understood personal tic-related experiences encouraged uncensored and honest discussions, and so respondents depicted online support communities as cathartic outlets in which reassurance, advice, and solidarity were plentiful. This may be in keeping with Cohen & Wills' [41] direct effect hypothesis: through sharing negative experiences to an online support community, others can provide support, thus rewarding interactions that plausibly shield people with TD (and their supporters) from maladaptive threat appraisals and associated psychosocial impairments [18,38,42].

The culmination of emotional and informational resources seemingly perpetuates supportive online support communities. Many people with TS and families experience rejection in society due to tic-related stigma [43], and exclusion can precipitate social isolation, causing many with TDs to strive for community acceptance [6]. In this study, integration into online support communities was described as a privilege and one which helped reduce the social exclusion experienced offline. Understandably, respondents were protective of the community tone and expressed strong attachments to particularly supportive online support communities. Desire to maintain inclusive communities encouraged many to adopt the supportive role themselves and actively induct newer, inexperienced members. As such, online support communities provide a premade community in which acceptance is not a major struggle.

Meeting other people with TD and comparable difficulties within a group therapy setting may be sufficient enough to improve QoL [44], and several respondents reported online support community use prompted reappraisals of personal circumstances that in turn facilitated positive psychological adjustment (eg, improved confidence, hope for the future, engaging in regular social activities). Smith et al [6] support this, advocating that acceptance of tics into personal identity precipitates quicker adaptation to long-term conditions. Perhaps offering insight into the mechanisms of these psychosocial changes, Cohen and Wills' [41] buffering hypothesis suggests that by illustrating what people with TD can achieve, supportive others (eg, caregivers) may reframe personal difficulties and help other online support community users from succumbing to maladaptive cognitions which endorse socially isolating behaviors. Additionally, several respondents talked about reappraising the stigma of self-blame and *punishment* they attributed to themselves for having tics. As well as other processes feeding into this reappraisal (eg, acceptance, improved confidence), another factor contributing to this could be through psychoeducational processes within online support communities, such as changing illness representations and new knowledge (eg, understanding more about the neurodevelopmental causes of TDs, the group mentality in online support communities in how they perceive and comprehend TDs). Furthermore, writing about personal experiences in health-specific dedicated online support communities allows users to authentically express themselves and may contribute to sense of empowerment in living with chronic conditions [45]. The findings also suggest online support community members become role models to others, whose adaptive behaviors and attitudes encourage others to participate in typical activities irrespective of tic-related impairments.

Despite the plethora of online support community benefits, participants were forthcoming about a number of limitations. Primarily, the suggestible nature of TDs perpetuated temporary onset of new tics and heightened tic severity upon exposure to graphic posts within online support communities. Similarly, vivid imagery in online support communities can be triggering for people who self-harm [23]. For people with TDs, this appears to be a *double-edged sword* in seeking out support: respondents have described the high value they place in communicating with similar peers, but it comes with risk of experiencing changes in tics (eg, increased frequency, new tics). Respondents in this study felt warnings preceding descriptions or videos of tics would be important in helping other users, and could be enforced by online support community moderators/administrators. Moreover, similar to Holbrey and Coulson [27], a number of members found reading others' negative TD-related experiences reinforced negative feelings and low mood. Further research is necessary to establish whether this drawback is more prevalent among those with depressive symptomology; however, at present, health care professionals should exert caution when recommending online support communities to people with TD and their families by informing them of the potential risks to their psychological well-being.

As similarly reported for other health issues [23,40,46], conflict between online support community members was a further drawback of engaging in online communities, and should perhaps caution parents of younger community members that online support communities for TD hold similar risks to other internet-based activities. The dominant conflict described by some respondents was resentment among adults with TD toward parents seeking advice regarding their child. Supporting Malli et al [18], adult respondents expressed frustration regarding minimal offline support for adults with TD, and so parental domination of online spaces was felt to further impede access to peer networks. Furthermore, parents' use of online support communities to discuss the negative impact of a child's TD endorsed insecurities among adults with TD and concerns regarding familial resentment of them. Having distinct online support communities for people with TDs and their caregivers could be an option to reduce this conflict, but at the same time, there is a balance in that caregivers may learn much from people who have TDs.

Strengths and Limitations

To our knowledge, this is the first piece of research to examine online support community use for TS/TDs and may provide useful recommendations for practice: online support communities can be proposed as one accessible solution—with cautions—to people with TD and their families otherwise unable to access social support. Moreover, the study positively contributes to a growing literary body pertaining to use of online support communities among those experiencing chronic long-term conditions.

There were several limitations. Although a diverse sample of 90 people of varying ages, genders, and mix of TD-related experiences from across multiple countries participated, they were mostly from Westernized countries and from participants who could respond in English. Future research may wish to explore non-English language online support communities to be more representative. Similarly, the majority of respondents being female was surprising given the higher prevalence of TDs in males [2], and so suggests the acquired sample was not necessarily representative of the prevalence of TD by gender. Furthermore, respondents had not all received a TD diagnosis—with several reporting suspected (but not confirmed) TD—alongside self-reported comorbidities. The nature of the study meant we were unable to confirm whether participants' suspected TD was diagnosable, and whether comorbidities had

been diagnosed too. Additionally, the sample was largely recruited through active online support communities: thus those with negative experiences who discontinued participation were not recruited, plausibly preventing the extent of online support community drawbacks from emerging. Although efforts were made to ask permission from moderators to post the study advert to several online support communities, ultimately 7 declined without explaining their reasons. It may be that patient-led groups wish to stay focused on patient-led discussion and do not want external organizations in their online spaces (eg, due to negative experiences with them; preferences to keep online groups for patients only). Finally, the nature of online survey methodology prevented follow-up questions, which may have provided greater insight and it remains unknown if other unique aspects of online support community use for TDs were not uncovered in this study.

Conclusion

Overall, online support communities were illustrated as an accessible support source for people with TDs and supportive others. Provision of informational and emotional support was delineated as the foundation for building online support communities, which facilitated a much appreciated sense of inclusion among users, and perpetuated improvements in psychosocial well-being and functioning. As such, despite their drawbacks, online support communities were largely considered hugely beneficial to those experiencing TDs and could be one resource that clinicians may recommend to socially isolated patients and families.

Conflicts of Interest

EBD is a named member of the ORBIT trial (ISRCTN70758207).

References

1. American Psychological Association. Neurodevelopmental Disorders. Diagnostic and Statistical Manual Of Mental Disorders (5th Ed). Arlington, VA: American Psychological Association; 2013.

2. Evans J, Seri S, Cavanna AE. The effects of Gilles de la Tourette syndrome and other chronic tic disorders on quality of life across the lifespan: a systematic review. Eur Child Adolesc Psychiatry 2016 Sep;25(9):939-948 [FREE Full text] [doi: 10.1007/s00787-016-0823-8] [Medline: 26880181]

3. Knight T, Steeves T, Day L, Lowerison M, Jette N, Pringsheim T. Prevalence of tic disorders: a systematic review and meta-analysis. Pediatr Neurol 2012 Aug;47(2):77-90. [doi: 10.1016/j.pediatrneurol.2012.05.002] [Medline: 22759682]

4. Bloch MH, Leckman JF. Clinical course of Tourette syndrome. J Psychosom Res 2009 Dec;67(6):497-501 [FREE Full text] [doi: 10.1016/j.jpsychores.2009.09.002] [Medline: 19913654]

5. Yadegar M, Guo S, Ricketts EJ, Zinner SH. Assessment and Management of Tic Disorders in Pediatric Primary Care Settings. Curr Dev Disord Rep 2019 Sep 7;6(3):159-172 [FREE Full text] [doi: 10.1007/s40474-019-00168-8] [Medline: 32467820]

6. Smith H, Fox JRE, Trayner P. The lived experiences of individuals with Tourette syndrome or tic disorders: a meta-synthesis of qualitative studies. Br J Psychol 2015 Nov;106(4):609-634. [doi: 10.1111/bjop.12118] [Medline: 25721405]

7. Cuenca J, Glazebrook C, Kendall T, Hedderly T, Heyman I, Jackson G, et al. Perceptions of treatment for tics among young people with Tourette syndrome and their parents: a mixed methods study. BMC Psychiatry 2015 Mar 11;15:46 [FREE Full text] [doi: 10.1186/s12888-015-0430-0] [Medline: 25879205]

8. Woods DW, Conelea CA, Himle MB. Behavior therapy for Tourette's disorder: Utilization in a community sample and an emerging area of practice for psychologists. Professional Psychology: Research and Practice 2010;41(6):518-525. [doi: 10.1037/a0021709]

9. Sharp AN, Singer HS. Standard, Complementary, and Future Treatment Options for Tics. Curr Dev Disord Rep 2018 Mar 22;5(2):101-107. [doi: 10.1007/s40474-018-0138-1]

10. Whittington C, Pennant M, Kendall T, Glazebrook C, Trayner P, Groom M, et al. Practitioner Review: Treatments for Tourette syndrome in children and young people - a systematic review. J Child Psychol Psychiatry 2016 Sep;57(9):988-1004. [doi: 10.1111/jcpp.12556] [Medline: 27132945]

11. Pringsheim T, Holler-Managan Y, Okun MS, Jankovic J, Piacentini J, Cavanna AE, et al. Comprehensive systematic review summary: Treatment of tics in people with Tourette syndrome and chronic tic disorders. Neurology 2019 May 06;92(19):907-915. [doi: 10.1212/wnl.0000000000007467]

XSL•FO

RenderX

12. Cravedi E, Deniau E, Giannitelli M, Xavier J, Hartmann A, Cohen D. Tourette syndrome and other neurodevelopmental disorders: a comprehensive review. Child Adolesc Psychiatry Ment Health 2017;11:59 [FREE Full text] [doi: 10.1186/s13034-017-0196-x] [Medline: 29225671]

13. Kumar A, Trescher W, Byler D. Tourette Syndrome and Comorbid Neuropsychiatric Conditions. Curr Dev Disord Rep 2016;3(4):217-221 [FREE Full text] [doi: 10.1007/s40474-016-0099-1] [Medline: 27891299]

14. Conelea CA, Woods DW, Zinner SH, Budman CL, Murphy TK, Scahill LD, et al. The impact of Tourette Syndrome in adults: results from the Tourette Syndrome impact survey. Community Ment Health J 2013 Feb;49(1):110-120. [doi: 10.1007/s10597-011-9465-y] [Medline: 22052430]

15. Wadman R, Glazebrook C, Beer C, Jackson GM. Difficulties experienced by young people with Tourette syndrome in secondary school: a mixed methods description of self, parent and staff perspectives. BMC Psychiatry 2016 Jan 20;16(1):14 [FREE Full text] [doi: 10.1186/s12888-016-0717-9] [Medline: 26792211]

16. Ludlow AK, Brown R, Schulz J. A qualitative exploration of the daily experiences and challenges faced by parents and caregivers of children with Tourette's syndrome. J Health Psychol 2018 Dec;23(14):1790-1799. [doi: 10.1177/1359105316669878] [Medline: 27682339]

17. Wadman R, Tischler V, Jackson GM. 'Everybody just thinks I'm weird': a qualitative exploration of the psychosocial experiences of adolescents with Tourette syndrome. Child Care Health Dev 2013 Nov 30;39(6):880-886. [doi: 10.1111/cch.12033] [Medline: 23363392]

18. Malli MA, Forrester-Jones R, Triantafyllopoulou P. "Tourette's Is a Lonely Place": an Interpretative Phenomenological Analysis of the Personal Experience and Identity of Adults with Tourette's Syndrome. J Dev Phys Disabil 2019 May 14;31(6):819-845. [doi: 10.1007/s10882-019-09676-2]

19. Reblin M, Uchino BN. Social and emotional support and its implication for health. Curr Opin Psychiatry 2008 Mar;21(2):201-205 [FREE Full text] [doi: 10.1097/YCO.0b013e3282f3ad89] [Medline: 18332671]

20. Goussé V, Czernecki V, Denis P, Stilgenbauer JL, Deniau E, Hartmann A. Impact of Perceived Stress, Anxiety-Depression and Social Support on Coping Strategies of Parents Having A Child With Gilles de la Tourette Syndrome. Arch Psychiatr Nurs 2016 Feb;30(1):109-113. [doi: 10.1016/j.apnu.2015.08.017] [Medline: 26804511]

21. O'Hare D, Eapen V, Grove R, Helmes E, McBain K, Reece J. Youth with Tourette syndrome: Parental perceptions and experiences in the Australian context. Australian Journal of Psychology 2016 Mar 21;69(1):48-57. [doi: 10.1111/ajpy.12111]

22. Smedley RM, Coulson NS. A practical guide to analysing online support forums. Qualitative Research in Psychology 2018 May 14:1-28. [doi: 10.1080/14780887.2018.1475532]

23. Coulson NS, Bullock E, Rodham K. Exploring the Therapeutic Affordances of Self-Harm Online Support Communities: An Online Survey of Members. JMIR Ment Health 2017 Oct 13;4(4):e44 [FREE Full text] [doi: 10.2196/mental.8084] [Medline: 29030324]

24. Prescott J, Hanley T, Ujhelyi K. Peer Communication in Online Mental Health Forums for Young People: Directional and Nondirectional Support. JMIR Ment Health 2017 Aug 02;4(3):e29 [FREE Full text] [doi: 10.2196/mental.6921] [Medline: 28768607]

25. Thoren EM, Metze B, Bührer C, Garten L. Online support for parents of preterm infants: a qualitative and content analysis of Facebook 'preemie' groups. Arch Dis Child Fetal Neonatal Ed 2013 Nov;98(6):F534-F538. [doi: 10.1136/archdischild-2012-303572] [Medline: 23759520]

26. Malinen S. Understanding user participation in online communities: A systematic literature review of empirical studies. Computers in Human Behavior 2015 May;46:228-238 [FREE Full text] [doi: 10.1016/j.chb.2015.01.004]

27. Holbrey S, Coulson NS. A qualitative investigation of the impact of peer to peer online support for women living with polycystic ovary syndrome. BMC Womens Health 2013 Dec 17;13(1):51 [FREE Full text] [doi: 10.1186/1472-6874-13-51] [Medline: 24341398]

28. Coulson NS. How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. JRSM Short Rep 2013;4(8):2042533313478004 [FREE Full text] [doi: 10.1177/2042533313478004] [Medline: 24040493]

29. Turner JA. Online Support Groups: The Good, the Bad, and the Motivated. Journal of Consumer Health on the Internet 2017 Mar 10;21(1):11-25. [doi: 10.1080/15398285.2017.1279930]

30. Lee M, Mu P, Wang W, Wang H. 'Living with tics': self-experience of adolescents with Tourette syndrome during peer interaction. J Clin Nurs 2016 Feb;25(3-4):463-471. [doi: 10.1111/jocn.13074] [Medline: 26818371]

31. Buchanan H, Coulson NS. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269. [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]

32. Coulson NS, Smedley R, Bostock S, Kyle SD, Gollancz R, Luik AI, et al. The Pros and Cons of Getting Engaged in an Online Social Community Embedded Within Digital Cognitive Behavioral Therapy for Insomnia: Survey Among Users. J Med Internet Res 2016;18(4):e88 [FREE Full text] [doi: 10.2196/jmir.5654] [Medline: 27113540]

33. Hall CL, Davies EB, Andrén P, Murphy T, Bennett S, Brown BJ, ORBIT Trial Team. Investigating a therapist-guided, parent-assisted remote digital behavioural intervention for tics in children and adolescents-'Online Remote Behavioural Intervention for Tics' (ORBIT) trial: protocol of an internal pilot study and single-blind randomised controlled trial. BMJ Open 2019 Jan 03;9(1):e027583. [doi: 10.1136/bmjopen-2018-027583] [Medline: 30610027]

34. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

35. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Self-Reported Differences in Empowerment Between Lurkers and Posters in Online Patient Support Groups. J Med Internet Res 2008;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]

36. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]

37. O'Hare D, Eapen V, Helmes E, McBain K, Reece J, Grove R. Factors impacting the quality of peer relationships of youth with Tourette's syndrome. BMC Psychol 2015 Sep 30;3(1):34 [FREE Full text] [doi: 10.1186/s40359-015-0090-3] [Medline: 26424471]

38. Lee M, Wang H, Chen C, Lee M. Social adjustment experiences of adolescents with Tourette syndrome. J Clin Nurs 2019 Jan;28(1-2):279-288. [doi: 10.1111/jocn.14564] [Medline: 29893437]

39. Coulson N, Shaw R. Nurturing health-related online support groups: Exploring the experiences of patient moderators. Computers in Human Behavior 2013 Jul;29(4):1695-1701 [FREE Full text] [doi: 10.1016/j.chb.2013.02.003]

40. Shoebotham A, Coulson NS. Therapeutic Affordances of Online Support Group Use in Women With Endometriosis. J Med Internet Res 2016 May 09;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]

41. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]

42. Cavanna AE. Gilles de la Tourette syndrome as a paradigmatic neuropsychiatric disorder. CNS Spectr 2018 Jun;23(3):213-218. [doi: 10.1017/S1092852918000834] [Medline: 29781408]

43. Eapen V, Cavanna AE, Robertson MM. Comorbidities, Social Impact, and Quality of Life in Tourette Syndrome. Front Psychiatry 2016;7:97 [FREE Full text] [doi: 10.3389/fpsyt.2016.00097] [Medline: 27375503]

44. Nissen JB, Kaergaard M, Laursen L, Parner E, Thomsen PH. Combined habit reversal training and exposure response prevention in a group setting compared to individual training: a randomized controlled clinical trial. Eur Child Adolesc Psychiatry 2019 Jan;28(1):57-68 [FREE Full text] [doi: 10.1007/s00787-018-1187-z] [Medline: 29956034]

45. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Computers in Human Behavior 2008 Sep;24(5):1867-1883. [doi: 10.1016/j.chb.2008.02.004]

46. Mo PKH, Coulson NS. Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. Int J Nurs Stud 2014 Jul;51(7):983-993. [doi: 10.1016/j.ijnurstu.2013.11.006] [Medline: 24342449]

Abbreviations

CBD: Cannabidiol-based products
CBIT: Comprehensive Behavioural Intervention for Tics
QoL: quality of life
TD(s): Tic disorder(s)
TS: Tourette syndrome

XSL•FO
RenderX

Original Paper

Expert Opinions on Web-Based Peer Education Interventions for Youth Sexual Health Promotion: Qualitative Study

Philippe Martin[1,2,3,4*], MSc; Corinne Alberti[1*], Prof Dr, PhD; Serge Gottot[4*], PhD; Aurelie Bourmaud[1*], MD, PhD; Elise de La Rochebrochard[2,3*], PhD

[1]Institut National de la Santé Et de la Recherche Médicale (INSERM), Epidémiologie clinique et évaluation économique appliquées aux populations vulnérables (ECEVE), Université de Paris, Paris, France

[2]Institut National d'Etudes Démographiques (INED), UR14 – Sexual and Reproductive Health and Rights, Aubervilliers, France

[3]Centre de Recherche en épidémiologie et Santé des Populations (CESP) - UMR U1018 (UVSQ/INSERM), Université Paris-Saclay, Université Paris-Sud, Villejuif, France

[4]GDID Santé, Paris, France

*all authors contributed equally

Corresponding Author:
Philippe Martin, MSc
Institut National de la Santé Et de la Recherche Médicale (INSERM)
Epidémiologie clinique et évaluation économique appliquées aux populations vulnérables (ECEVE)
Université de Paris
10 Avenue de Verdun
Paris, 75010
France
Phone: 33 676606491
Email: philippe.martin@inserm.fr

Abstract

Background: Participatory education, in the form of peer education, may be an effective way to promote youth sexual health. With the advent of the internet, web-based interventions have potential as an attractive new tool for sexual health promotion by peers.

Objective: The aim of this study was to evaluate professional experts' opinions on the perspectives for web-based participatory interventions to promote sexual health by peers and among young people.

Methods: Semistructured interviews were carried out with 20 experts (stakeholders in direct contact with young people, researchers, and institutional actors) specializing in sexual health, health promotion, peer education, youth, internet, and social media. After coding with N'Vivo, data were subjected to qualitative thematic analysis.

Results: The majority of experts (18/20, 90%) found this kind of intervention to be attractive, but highlighted the necessary conditions, risks, and limitations attached to developing an acceptable peer intervention on the internet for sexual health promotion among young people. Five main themes were identified: (1) an internet intervention; (2) sexual health; (3) internet skills, and uses and the need for moderation; (4) multifaceted peers; and (5) minority peers. In the absence of youth interest for institutional messages, the experts highlighted the attractive participatory features of web-based interventions and the need for geolocalized resources. However, they also warned of the limitations associated with the possibility of integrating peers into education: peers should not be mere messengers, and should remain peers so as not to be outsiders to the target group. Experts highlighted concrete proposals to design an online participatory peer intervention, including the process of peer implication, online features in the intervention, and key points for conception and evaluation.

Conclusions: The experts agreed that web-based participatory interventions for youth sexual health promotion must be tailored to needs, uses, and preferences. This type of action requires youth involvement framed in an inclusive and holistic sexual health approach. Peer education can be implemented via the internet, but the design of the intervention also requires not being overly institutional in nature. Involving young people in their own education in an interactive, safe online space has the potential to develop their empowerment and to foster long-term positive behaviors, especially in the area of sexual health.

(J Med Internet Res 2020;22(11):e18650) doi:10.2196/18650

XSL·FO
RenderX

Introduction

UNESCO (United Nations Educational, Scientific and Cultural Organization) emphasizes that adolescence is the time to develop healthy habits and lifestyles related to sexual health, when individuals are exploring their sexuality and establishing interpersonal relationships [1]. Young people have emerging questions about sexual health, which go beyond disease and integrate emotional dimensions [2,3]. Major sexual and reproductive health issues affect young people, including puberty, sexually transmitted infections [4], unwanted pregnancy [5], and sexual violence, but information/communication technologies also have an influence on sexual behavior [1].

To address these issues, top-down sexual health education has been developed. However, learning is not limited to receiving and processing information: young people learn best when they are allowed to build their own understanding of information [1]. Among young people, interactive models that promote social interaction and exchanges of experiences could be effective in acquiring this knowledge and developing positive health behavior over the long term [6]. In this process, peer education corresponds to an educational approach that uses peers (sharing the same age, social context, function, education, or experience) to provide information and to promote certain types of behaviors and values [7,8]. Through informal social learning, peers appear as resources offering support and sharing similar experiences. Health actors have tried to advance peer education to become a formal process in public policy and practice, particularly with respect to youth sexual health [9-12].

In the past, peer education programs have been implemented in physical life, especially for HIV prevention and sexual health education [9,13,14], with different peer-led group sessions [9]. Past experiences have highlighted the value of a comfortable and user-friendly space to exchange information and perspectives about sexuality, with peer educators facilitating youth engagement [15]. A recent review of college campus peer interventions found positive improvements in knowledge and behaviors such as condom use and HIV testing [16].

With the advent of new technologies, the internet offers wide access to health information, particularly in sexual health [3], with benefits of interactivity and personalization of information [17]. Social media offer users the ability to generate, share, and receive information through multidirectional exchanges, which can transcend geographical boundaries and provide anonymity in discussing intimate topics [18,19]. Young people can join online communities to benefit from social support and find responses to their concerns. The online media have potential to offer great opportunities for peer education interventions in sexual health for young people.

Despite these recognized benefits, there is little research evaluating participatory interventions on the internet and by peers in youth sexual health. One intervention study explored the feasibility of peer education among adults focusing on men who have sex with men (MSM) for HIV prevention, with the training of leaders in a Facebook group [20]. Another study designed a social media "peer-led" intervention in a Facebook group for safer sex practices among young people [21]. Although the feasibility of these interventions has been demonstrated, much work remains to be done to determine whether this educational model is applicable and effective, and if it is complementary to traditional top-down systems. In particular, further research is needed to explore the educational and practical potential of these interventions, examining inherent risks and limitations. Feedback from experts (in the fields of internet, youth, and sexual education) should make it possible to address several of the key methodological issues.

Community-based participatory research brings together partners (actors in the field, designers and researchers, target audience) with different skills, knowledge, and expertise to address complex problems, including experienced professionals [22]. Collecting their views based on their experiences should inform the design of in-depth analyses, particularly for the development of peer education interventions on youth sexual health [15]. Currently, there are no data available on relevant stakeholders' opinions and experiences on peer education, sexual health, youth, and the internet. Therefore, there is a need for informative research to fill these gaps, and to develop this new kind of intervention effectively.

Accordingly, the aim of this study was to evaluate expert opinions and to collect advice on web-based participatory interventions to promote young people's sexual health through peers, and to study the inherent risks and limitations.

Methods

Design

Our methodology followed the Consolidated criteria for reporting qualitative research (COREQ) checklist for writing and reading qualitative research reports [23] (see Multimedia Appendix 1).

Participant Recruitment

We identified 37 French experts on the subject from existing publications in which they were named as authors, as well as reports and health promotion programs in which they were credited as editors and managers. These experts had different functions in diverse fields of expertise, including sexual health, young people's health, connected health, peer education, and program methodology. They all had concrete experience on the subject and had expertise related to some or all of the study topic. Some of the experts were in charge of websites promoting the sexual health of young people, were developing educational programs or national sexual health guidelines, or had studied the sexual health of young people, particularly through the internet. We invited these experts to participate in the study by email. Of the 37 identified experts, 20 agreed to participate. The characteristics of the participant experts are presented in Table 1 (see Multimedia Appendix 2 for more details).

Table 1. Expert characteristics (N=20).

Characteristic	Participants, N
Gender	
Female	13
Male	7
Organization type	
Association	5
Public institute for prevention/health promotion	5
Public hospital	4
National education	2
Public research institute	2
Public scientific/cultural/professional institution	1
Government agency	1
Occupation	
Professor of health (public health, health promotion, gynecology)	3
Nurses and midwives specializing in sexual health	3
Clinical psychologists/youth psychotherapists	2
Heads of prevention department/prevention project manager	2
Sociologist experts on online youth sexuality	2
College teacher (science and sexuality education)	1
Regional health education program coordinator	1
Prevention facilitator; specialized educator	1
Advisor in social and family economy	1
Social marketing expertise manager	1
Documentation and information officer	1
Epidemiologist	1
Social worker	1
Specializations	
Youth health	18
Sexual health	16
Education/prevention/promotion	12
Peer education	9
Internet and social media	5
Region	
Paris region	16
Rest of France	4

Interview Process

A researcher who is a graduate in public health (PhD candidate) and trained in interview techniques conducted the semistructured interviews with professional experts. Each interview began with a presentation of our research subject and key associated concepts. The interview guide was not constructed based on a specific theoretical framework. Open-ended questions organized in a convergent manner were used to possibly prompt the interviewee on a subject. The content of the questions was based on factual information–seeking only. The semistructured interviews then followed the guide (see Multimedia Appendix 3) with adaptation depending on the interviewee's experience. The interview guide was structured in four main sections: (1) features of experts, (2) youth sexuality concerns, (3) mechanisms for seeking information or exchange of experiences, and (4) opinions and experiences on web-based interventions and participatory features as peer education.

Interviews were audio-recorded after having obtained the agreement of the interviewed expert. The interviews lasted between 45 and 141 minutes with an average of 63 minutes. Most of the interviews were with individuals; only one interview was held with two experts working together. Interviews were carried out by telephone or at the experts' place of practice, allowing the expert to remain in his or her working environment to complete the interview.

Regulatory and Ethical Aspects

The study obtained a favorable opinion (no. 18-515) from the Institut National de la Santé Et de la Recherche Médicale (INSERM) Ethics Evaluation Committee (IRB0000003888) and was reported to the INSERM Data Protection Officer. Identifying information was anonymized in the transcripts. Expressed consent was given orally at the beginning of the interview, as was the authorization to record.

Analysis

The interviewer transcribed and coded the digitized interviews verbatim and the notes taken during the interviews using N'Vivo 10 software. Using an inductive theme identification process to generate codes, an analytical framework was created. To ensure the validity of the results, the thematic analysis was carried out by two authors (PM, ELR), who followed the recommended phases and steps for the development of the themes in terms of qualitative content and thematic analysis [24]: Initialization, Construction, Rectification, and Finalization.

Results

Main Themes and Subthemes

Experts' opinions were structured according to five main themes: (1) internet intervention; (2) sexual health; (3) internet skills and uses, and need for moderation; (4) multifaceted peers; and (5) minority peers. These themes are complementary, connected, and should be considered together for the development of participatory internet-based and sexual health promotion peer interventions. The themes and subthemes are presented in Textbox 1 (also see Multimedia Appendix 4 for the most significant quotes used to develop these themes).

Textbox 1. Key themes and subthemes identified in the interviews.

Theme 1: Internet intervention

- Complementarity with existing offline sexual tools

- Secure, valid, and credible content

- Online personalized, interactive, and participatory features

- Adapting to rapid obsolescence of preferred media

- Social marketing to understand uses and preferences

Theme 2: Sexual health

- Importance of sexual health for young people

- Evolving sexual health concerns and issues

Theme 3: Internet skills and uses, and the need for moderation

- Heterogeneous internet use

- Diversity of skills in online information–seeking

- Need for online anonymity for sexual issues

- Risks of surfing the internet and social media

- Moderation of online social interactions

- Nonreceptivity of institutional messages

Theme 4: Multifaceted peers

- Importance of peer group for young people

- Peer education concept

- Notion of peers

- Peers' involvement

- Limits of young people as "peer educators"

Theme 5: Minority peers

- Online peer group dynamic

- Need to find peers outside the neighborhood

- Inclusiveness in health promotion interventions

- Risk of stigmatization and discrimination

- Self-rejection as a determinant of participation

- Collaboration with specialized organizations

Internet Intervention

In view of young people's daily internet use, the majority of experts (18/20, 90%) believe that web-based participatory interventions to promote youth sexual health are attractive, with essential prerequisites. One expert found the intervention to be attractive without suggesting any limit. Another did not find the intervention to be attractive (owing to too many limitations).

Complementarity With Existing Offline Sexual Tools

A program coordinator insisted that a web-based action must be complementary to existing offline sexual health tools and actions (ie, the internet does not replace human contact; expert S5). Another highlighted the need to "digitize" existing health promotion techniques (S17).

Secure, Valid, and Credible Content

An expert in direct contact with youth indicated that to address young people through a web-based intervention, she will need to be sure that it is well embedded and disseminates valid and credible contents (S13).

Online Personalized, Interactive, and Participatory Features

Three experts (S1, S15, and S17) insisted on the importance of using internet tools to develop participatory actions, to go beyond observing or receiving information. Two experts (S1 and S17) mentioned online participatory features as attractive components, including shared construction of knowledge, the possibility of including influencers (role models), serious games, and chatbots (artificial intelligence). Two other experts (S8 and

S19) considered that young people may have needs for health services offline beyond the internet, particularly in sexual health (eg, access to abortion, protection from violence). They also suggested providing geolocalized resources to allow for a personalized response, according to participant location.

> *For an internet site, it is important to be able to say, for a local area, where I can get more information, where can I get condoms? It must be locally sited and rooted in an area.* [S8]

Adapting to Rapid Obsolescence of Preferred Media

One expert (S5) warned that young people's preferred online media sources are evolving quickly. Therefore, it is necessary to adapt actions to the evolution of internet uses and the rapid obsolescence of these preferred media.

> *It will already be obsolete, and will no longer correspond to their favorite network. This is something that changes fast, so for it to be set up and be effective, we have to be reactive.* [S5]

Two experts (S5 and S15) also proposed to use multiple online media sources in parallel (interconnected), to observe preferences, and to be as close as possible to media uses.

Social Marketing to Understand Uses and Preferences

Moreover, a communications specialist pointed out the need for a social marketing strategy to understand young people's internet usage and to adapt to their preferences in implementing actions.

> *How to attract young people? It has to be a brand, there must be a marketing strategy, one has to think about several different sites. What we are studying is peer education, which is quite well known. But one just has to get on terms with different internet sites and know how to make them work.* [S15]

Sexual Health

Importance of Sexual Health for Young People

Based on her research, a sociologist expert on young people's internet usage and sexuality (S19) explained that among other health issues, those related to sexual health appeared to be the most important among young people. Three experts (S1, S9, and S11) emphasized the emotional dimensions of sexual health: love and sentimentality, as well as sexual relationships are central for young people. In this sense, three experts (S1, S2, and S4) also recalled the importance of a holistic approach to sexual health action, treating it in a global way and going beyond the prevention of risks.

Evolving Sexual Health Concerns and Issues

A sociologist expert on young people's sexuality as manifested online (S16) emphasized the importance of taking into account, when planning actions, the evolution of a young person's sexuality concerns, especially following their sexual debut. He recommended recognizing that internet use evolves with life trajectories: "Internet usage is linked to one's situation, and depends on age and on the changing concerns implicated in one's emotional and gender relations" (S16).

Internet Skills and Uses, and the Need for Moderation

The majority of the experts (19/20, 95%) discussed young people's internet usage as a route to address sexual issues (preferred media, use of social media to interact and search information). They emphasized the heterogeneity of the young population.

Heterogeneous Internet Use

One sociologist (S16) explained that exploring the internet may be a solitary activity at first when seeking to understand sexuality; in this phase, young people do not necessarily want to interact with others. However, they may ask questions on online search engines to find information and observe the exchanges in forum discussions (S3, S16).

> *Early adolescence, before the first sexual relations, is often a time of very solitary exploration of the internet, it is not social networks that are the most important. But at this stage young people do follow forums [discussions online].* [S16]

Diversity of Skills in Online Information–Seeking

Based on her experience, another expert (S19) insisted that heterogeneous internet usage for information retrieval must be considered. There is a diversity of backgrounds and different skills in research, analysis, and critical thinking concerning online information:

> *Young people from better-off backgrounds, who have the greatest inclination and also the most social, educational, economic, and cultural capital, will be the ones who will make the most use of the different resources the internet has to offer.* [S19]

On the same subject, four experts in direct contact with young people (S2, S3, S9, and S11) made the point that they are eager to find answers online and are not always critical as to the reliability of the source. However, two experts in direct contact with young people (S6) and in sociology (S16) considered young people to be sufficiently competent to screen online information.

> *Young people have quite a strong tendency to resort to the internet, and their grasp of technology enables them to know the difference between a site which gives valid advice and one which looks untrustworthy.* [S16]

Need for Online Anonymity for Sexual Issues

Moreover, given the intimate and personal nature of the topic of sexual health, three experts (S5, S6, and S19) mentioned the advantage of online anonymity as protection for those needing to ask questions or seek information. In practical terms, one explained the advantage of this: "They have anonymity already, so they will be able to ask their questions more easily than face to face, from behind their screen…without embarrassment or fear of judgment by their peers." (S5)

One expert (S19) complemented this notion by explaining that young people generally leave their usual media sources that can identify them to go to other sites so as to have anonymity for questions on sexuality. However, two experts (S6 and S9) also pointed out that anonymity could have drawbacks, with the risk

XSL•FO

RenderX

of cyberbullying: "It can also be a protection for the most abusive or malicious, the internet 'trolls'." (S6)

Risks of Surfing the Internet and Social Media

Five experts in direct contact with young people (S2, S6, S9, S10, and S13) were those who gave the strongest warnings about online risks such as access to unreliable and invalid information (S3, S9), exposure of bodies by "nudes" (S2, S3, S10), or access to pornography (S2, S10).

> In schools and colleges there are problems to be dealt with which arise on social media, on Snapchat. We have had quite a lot of trouble with photos where girls are posing on social network sites. [S3]

Moderation of Online Social Interactions

To address these risks, the majority of experts (15/20, 75%) expressed the view that online moderation is necessary, notably in web-based actions allowing social interactions. Even in a peer education intervention, educators should not disengage from their adult role (S7, S8). This moderation must make it possible to reduce hurtful acts and limit false information (S2, S3):

> The moderator must be really good, so that as soon as there is a false statement, or one that is hurtful or insulting, the moderator intervenes. There would need to be a super-present moderator. [S3]

Nevertheless, two experts (S7 and S16) thought that young people might consider this moderation as imposed from the outside, thereby losing the desired "between young people" aspect.

Nonreceptivity of Institutional Messages

More generally, three experts (S6, S7, and S16) noted that young people are not attracted or receptive to traditional prevention actions and messages developed by institutions. Some experts (S5, S11, S15) pointed out that existing sexological tools (school interventions, websites) are effective. In contrast, two experts (S7 and S16) were critical of young people's perception of formal actions, which are considered to be not effective or too institutional (S7).

To address the nonreceptivity to institutional messages, two experts (S6 and S17) insisted on the need to involve young people in project reflection. Moreover, a teacher (S3) explained the importance of "peers" for the acceptance and integration of the information received: "We know that studies show that when information and knowledge are offered by a member of a peer group, it is better accepted and retained than when it is provided by the teacher."

Multifaceted Peers

Importance of Peer Group for Young People

For a sociologist expert on young people's sexuality online (S19), the peer group corresponds to one of the new spheres of socialization where peers will be chosen and will take up a lot of space in the lives of young people. Peers could then intervene in education.

Peer Education Concept

Among experts who had experience in peer education (9/20, 45%), the majority (S1, S6, S7, S8, S15, and S19) considered peer education as part of an approach that involves youth participation in action. For another expert (S4), peer education was described as a discussion time between young people. One expert insisted on the importance of not considering young people only as "action users," but rather including them in all stages of action development, not only in design:

> This is a group of young people who self-select to set up a project which can be designed and made available to young people who are like themselves. [….] They are the ones who will take the initiative, and will be involved in the design, the implementation, and the evaluation. [S6]

Notion of Peers

Six experts (S4, S6, S7, S8, S15, and S19) questioned the idea of "youth" being peers and the notion of "peers." They insisted that peers must recognize and consider themselves as peers (S6, S7, S8, and S19). One (S6) advocated letting the group of peers form themselves within the action, without institutional involvement. One (S19) indicated that it is complicated to simply consider the "age" characteristic as the gateway. For one expert, young people are all peers and she considered this as a limitation: "It just means young people talking to young people, so they are 'pseudo-pairs'." (S4)

For others (S6, S8), it seemed more pertinent to define the notion of "peers" in terms of similarity in experiences or concerns, beyond the criterion of age:

> "Similar" doesn't mean in terms of gender or skin color, but similar in terms of daily realities of life. They will have unity in terms of place, geographical space, and age. [S6]

Peers' Involvement

Two experts (S10 and S16) addressed the issue of young people's involvement in web-based peer education. Some could be leaders and others more passive: "Some young people will immediately want to position themselves as leaders within the group, and others will prefer to come and look and say nothing." (S10)

Three experts (S6, S7, and S8) identified different peer functions: moderators, educators, and receptors. One (S6) believed that peers could achieve online moderation: they are vigilant and autonomous for self-regulation. He suggested identifying "peer moderators" when the group is formed, based on peer involvement.

Limits of Young People as "Peer Educators"

In this sense, some young people could be selected to be "peer educators" (with young people taking over the action). However, three experts (S6, S7, and S8) highlighted the limitation of training peers to become "educators," considering that they would only become institutional messengers (S7, S8) or reproduce the same effect as the prevention facilitators (S6).

If the peers have been formatted by the institutions, they will become outsiders to the group. The group will quickly realize that there is an institution behind them, and they will keep away…because as soon as adolescents are transformed into health educators, they are no longer adolescents, they are spokespersons for the adults. They are spokespersons for approved messages. That's what I call parrots, the faithful repeaters of adult speech. [S7]

Yet, for one expert (S8), peer educators can facilitate close relationships, and people will believe information from peer educators because they are trained. For this expert, such peer educators may also have an effect on their social environment beyond their peers.

Experts were then divided between the need to let young people take over the action, with the right to their imperfections (S7), and the need for institutions to moderate and validate information and exchanges.

Minority Peers

Online Peer Group Dynamic

Based on his experience, a sociologist expert on young people's sexuality online (S16) explained that online peers are mainly the same people as physical peers but are also those engaged with for online interactions. Online social life is not generally separated from daily physical life (S19), with one exception:

For adolescents, it is rare to have a group of friends online which is completely different or much larger than one's physical group of friends, but there is one exception to this which I think is important, and that is the case of sexual minorities. [S16]

Need to Find Peers Beyond the Neighborhood

Some specificities could lead people to search out peers in online areas (S16). Sexual minority populations have specific needs, including finding peers on the internet and far from their immediate geographical area (S10, S16, and S19). In particular, lesbian, gay, bisexual, transsexual (LGBT) peers would be present, but outside of the immediate environment. For these young people, "online peers" are then part of their real lives and are not to be considered as "virtual" (S19): "In LGBT contexts, this is something we often find: accessing the internet to get in touch with a network which can't be located in certain geographical areas." (S19)

Inclusiveness in Health Promotion Interventions

To adapt the action to specificities, three experts (S15, S18, and S19) mentioned inclusiveness issues to be considered to take vulnerable populations into account (eg, LGBT, people with disabilities, overweight, or deaf). For one expert (S15),

representing minority populations in mainstream communication actions is a way to be inclusive. Another (S19) addressed the issue of "ourselves" and the need to form more specific subgroups: "We need to be inclusive in all our statements, at the same time it is good to form subgroups and to offer services which also correspond to within-group expectations." (S19)

The difficulty of being inclusive was underlined, as it requires significant material, human, and financial resources (S18) to have an intervention that speaks to all (S15).

If we want to be inclusive, then young white heterosexual girls will have to come up against queer people, young men who have sex with men, and come to terms with different life experiences. [S15]

Risk of Stigmatization and Discrimination

Moreover, one sexual health communications expert (S15) explained that risks of discrimination and stigmatization of specific audiences can occur in an all-audience activity. Minority people could perceive stigmatization and feel excluded, looked down upon, or treated differently.

For HIV, the recommendations are a bit different. For example, for a heterosexual person, testing is needed at some point during one's lifetime, whereas for MSM it has to be 4 times per year. So there one has to work carefully, because there begins to be a risk of stigmatization. [S15]

Self-Rejection as a Determinant of Participation

One expert (S15) raised the problem of self-rejection affecting participation in a health program. A person who does not accept themselves will not want to be part of a peer group and take part in an action addressing issues of sexuality.

In the case of young MSM, some will not be at all willing to come near a peer-led health education program in which there is a risk of even raising the idea that desire for other men exists. There may be a kind of self-rejection. [S15]

Collaboration With Specialized Organizations

To address discrimination and self-rejection, one expert (S15) raised the importance of collaborating with specialized associations/organizations to take into account points of view and specificities. These organizations could intervene to moderate online peer exchanges (S19).

Proposals Derived From the Themes

From the thematic analysis, several proposals could be drawn out, which are presented in Textbox 2, that may be of direct use in designing an online participatory intervention for peer sexual health promotion.

Textbox 2. Advice from experts for intervention conception and evaluation.

Domain 1: Peer intervention for sexual health promotion

- Conceptualize in advance what is behind the terms "peer," "peer-led," and "peer education"

- Complement existing online and offline educational practices by listing existing sexological tools

- Avoid institutional formatting of trained peer educators, as peers should remain peers. Peer leaders can be identified within an already formed community and then recruited as peer educators

- Define a framework for the involvement (interaction and participation) of young people

- Design a plan for facilitating and moderating peer exchanges

- Take a holistic approach to sexual and reproductive health, going beyond risk

- Identify in advance the specific needs of the populations

- Develop an inclusive, nonheteronormative approach that avoids stigmatization/discrimination of specific minority populations

- Involve local actors and associations to be as close as possible to the expectations of young people (with ability to provide answers in the moderation of exchanges)

Domain 2: Internet support

- Develop a secure online environment that fosters "self-confidence"

- Use online media that allow horizontal transmission of information (peer education), rather than top-down information systems

- Develop a brand image that is not institutional and that allows young people to recognize it in the online environment (use of social marketing techniques)

- Offer participatory (games, quizzes) and interactive (discussion forum, chatbot, questions and answers, possibility to contact a professional) features

- Integrate young people's favorite influencers and online characters, and the possibility of interacting with them

- Propose individual spaces (messaging, information folders) within the intervention to take into account the needs of solitary exploration

- Ensure anonymity of participants to encourage youth participation in intimate issues

- Grasp "youth culture" to be as close as possible to digital and online uses

- Be responsive in understanding the preferred tools and their use in an intervention

- Propose an interconnection of online media (website and social networks) to retain young people, integrate them into daily use, and observe preferences

- Provide a geolocalized response for access to sexual and reproductive health services or to meet with resource professionals offline and close to people's homes

Domain 3: Conception and evaluation

- Use the community-based participatory research model to involve all stakeholders at all stages of the project (bridge between research, field, and realities)

- Ensure the diversity of the peer group involved in the design and facilitation of the program to move beyond heterogeneity and integrate/represent minority populations (LGBTQ, deaf, disabled, overweight)

- Design a theoretical model to evaluate the effect of the intervention on:

  - Determinants of behavior change: knowledge, attitudes, literacy level, behaviors

  - Measure the effect of collective determinants on each of the determinants of behavior change: effects of online social interactions, perceived online social support, and online social capital

  - Analyze online peer social networks

- Define an operational framework for the online intervention, to define in advance the process indicators to be evaluated:

  - Journey within the youth intervention: those who are active or lurkers within the intervention

  - Most used features and tools

  - Number of visits, interaction and participation rates in the proposed activities (to be linked to the effectiveness evaluation)

Discussion

Principal Findings

This study is the first to highlight experts' opinions on key points and requisites for developing attractive and acceptable peer interventions on the internet for youth sexual health promotion. Experts considered this kind of intervention to be attractive, but warned of inherent risks and limitations. The experts interviewed provided very concrete and useful advice for developing web-based interventions for peer education, specifically in the field of sexual health.

One of the strengths of this work is that the experts did not confine themselves in the very purpose of arguing their "professional" point of view. By contrast, they envisioned the intervention by considering all the other actors involved, especially young people, with the help of their field experience. The themes addressed are therefore very broad, allowing the intervention to be thought of in its different dimensions with a global vision. This positioning of the experts is a very good indication of feasibility for the intervention because it shows that the experts project themselves in a global approach.

These results raise the following points for discussion: (i) there is a need to understand online uses and risks to take advantage of the internet; (ii) if peers are integrated into participatory education, they must recognize themselves as peers and must be selected for other characteristics than merely age; (iii) the notion of peers and specific audiences still needs to be understood to be inclusive in the web-based intervention.

Understanding Online Uses and Risks to Take Advantage of the Internet

Young people are daily users of the internet, but the experts stressed their heterogeneous skills in seeking information about sexuality, depending on sexual development. To adapt a web-based intervention for sexual health promotion, they highlighted participatory features but also underlined the need to provide geolocalized resources. They also emphasized the need to moderate content and exchanges. Understanding users by using social marketing and managing risks should make it possible to offer attractive, secure, and adapted interventions.

Grasping and understanding "youth culture" would make it possible to adapt health promotion actions to be as close as possible to users. Social marketing enables effective educational programs to be developed based on scientific knowledge and good communication [25]. To be reactive when faced with evolving preferences, it is advisable to design an interconnection of different media sources (eg, websites, apps, social network sites) and to offer attractive components (eg, games, individual spaces, forum discussions, or contact with influencers [26]). Important factors include the structure of the campaign and the content of the message (imaginative, fun, accessible, noninstitutionalized brand image, and engaging) [27].

However, young people may have difficulties in finding locally relevant information on health services [28]: needs may be expressed offline, and young users face many barriers of confidentiality, cost, and access to health services [29]. One solution is to adapt to target population environments [30] by providing geolocated information about access to offline services close to home. One example of this approach is the Australian organization PASH [31], which offers available local resources, including on-call resources, for personalized information.

To cope with online risks (eg, cyberbullying, pornography, body exposure), moderation should control interactions and abusive content, and provide a secure space that maintains the quality of information [32], while leaving space for user-generated content. For example, individuals may report a greater intention to participate in an online community that shows signs of moderation [33], but not if this moderation is involved too early [32]. A moderator should have an engaged and an interactive presence, and should be designed to generate new interests, provide discussion material, and respond to user requests [32]. The paradox is then to enable this moderation while at the same time allowing young people to be fully involved in the action.

Integrate Peers in Participatory Sexual Health Promotion

Facing the lack of focus on institutional messages, experts recognized the value of peer integration in web-based interventions. However, when discussing this possibility, many experts warned of limitations: peers should not be mere messengers and should remain peers so as not to be seen as outsiders by the group. The challenge then lies in young people taking ownership of their own education. One solution evoked was to identify active peers who can take the role of peer leaders once the online group is formed. Some experts thought that young people should be included at all stages of the project.

For sexuality education using digital media, it is recommended to use a variety of interactive methods and nonformal settings [1]. The appeal of peer sexuality education is that it has always existed on an informal basis, with young people sharing information with each other, including personal experiences [34]. From an action perspective, young people must be allowed to identify themselves as peers, particularly through the formation of subgroups within the intervention. "Peer" education can appear at this time, since the feeling of being "among peers" influences participation.

From a peer-education perspective, we first need to define the degree of involvement of peers in different steps of the project so that it remains a "peer-led" intervention. Next, it is necessary to allow them, through informational or experiential exchanges, to develop their knowledge and skills (peer education) [8,9]. This is also about justifying the inclusion of peers [35]. Many online interventions mobilize peer interactions to promote sexual health [18,36-41]; thus, a framework should make it possible to define the types of interactions between peers and to clarify the knowledge or skills to be developed. Based on the attractiveness of social interactions, the challenge is then to move from this informal mode to formal and conceptualized action led by institutions, particularly on the internet. It is also important to conceptualize "peers" and to define the common characteristics required for peers to be considered.

XSL•FO

RenderX

Understanding the Notion of Peers and Specific Audiences to Achieve Inclusivity

Experts do not always consider peers in the same way, with some prioritizing the "similar" aspect of a common level of life experiences, and others age.

Concretely, it would be interesting to consider the peer group as a nonfrozen and shifting network, built on common characteristics or paths, sharing not only an age characteristic but also sexual orientation, gender, health problem, or life experience. When considering "peer" education, we must define what this means, and when to differentiate "young people" and "peers." "Adolescent" peer groups can be characterized by a high degree of social solidarity and a code that contrasts with adult values [7]. Nevertheless, peers can find each other more easily by way of experience. For example, the transmission of older people's experiences of sexuality can sometimes establish a stronger link between peers of similar sexual orientation.

Considering this notion of peers, some young LGBT people, or those with disabilities or other characteristics, may feel excluded or unaffected by a general public action. For example, heteronormativity creates a sense of invisibility, invalidation, and marginalization for people of gender and sexual diversity [42]. Involving local actors or associations in actions permits an inclusive approach, to be close to specificities. Recommendations for inclusive research include using culturally appropriate language, not assuming that participants are heterosexual or that certain behaviors are "normal," and being aware of one's prejudices and knowledge limitations as a researcher [43]. In addition, community-based participatory research should integrate "all" target audiences into action management to elicit an appropriate response [22].

Strengths and Limitations

The strength of our study lies in the fact that it is the first to analyze expert opinions on the potential web-based interventions for peer-to-peer promotion of youth sexual health. These opinions thus complement the results of the few existing studies showing the feasibility of this type of intervention. Beyond the practical aspects and first demonstrations of effectiveness, previous studies such as the Harnessing Online Peer Education

(HOPE) study and peer-led, social media–delivered interventions have highlighted the major aspect of involvement of peer leaders [20,21], which we temper with our results (ie, peers must remain peers).

Moreover, our study proposes guidelines to design and implement this kind of intervention. The diversity of the experts interviewed makes it possible to obtain opinions from institutional, research, and field professionals. The analysis has made it possible to take into account expectations, realities, and obstacles on the sexual health, youth, education, and internet aspects.

One of our limitations is that we were unable to find researchers who had evaluated this type of intervention, since it has not yet been developed and evaluated. This would certainly have provided new methodological tools essential for action research. Despite this, we were able to interview developers of online sexual health content for young people who also worked on peer education. However, we did not interview peer educators who have participated in peer education programs and could have provided more information about this specific topic. This is an inherent limitation of our recruitment methodology. We could also have asked young people about their needs, expectations, and attractions for this type of intervention. This should be included when developing programs of action through participatory research.

Conclusion

Experts expressed the view that web-based participatory interventions for youth sexual health promotion by peers must be tailored to sexual health needs (information-seeking, socialization, services), the evolution of internet uses, and preferences in terms of participatory features. This type of action requires youth involvement in an inclusive and holistic sexual health approach. Peer education can be implemented on the internet, but the quality of the intervention also relies on not making the intervention too institutional. Involving young people in their own education in an interactive, safe online space then has the potential to develop their empowerment and long-term positive behaviors, especially in the area of sexual health.

Authors' Contributions

PM was responsible for all stages of the project, ELR, CA, SG, and AB participated in the design, analysis, and writing of the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO

RenderX

Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist.
[DOCX File , 15 KB - jmir_v22i11e18650_app1.docx]

Multimedia Appendix 2
Detailed table of experts' characteristics.
[DOCX File , 16 KB - jmir_v22i11e18650_app2.docx]

Multimedia Appendix 3
Interview guide.
[DOCX File , 15 KB - jmir_v22i11e18650_app3.docx]

Multimedia Appendix 4
Detailed themes and key quotes from the qualitative analysis.
[DOCX File , 30 KB - jmir_v22i11e18650_app4.docx]

References

1.   UNESCO. International Technical Guidance On Sexuality Education: An evidence-informed approach. Paris, France: UNESCO Publishing; 2018:142.
2.   Sexual health. World Health Organization. 2019. URL: http://www.who.int/topics/sexual_health/en/ [accessed 2019-11-21]
3.   Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
4.   Dehne K, Riedner G. Sexually transmitted infections among adolescents. World Health Organization. URL: https://www.who.int/maternal_child_adolescent/documents/9241562889/en/ [accessed 2019-05-22]
5.   Kar SK, Choudhury A, Singh AP. Understanding normal development of adolescent sexuality: A bumpy ride. J Hum Reprod Sci 2015;8(2):70-74 [FREE Full text] [doi: 10.4103/0974-1208.158594] [Medline: 26157296]
6.   Viero VDSF, Farias JD, Ferraz F, Simões P, Martins J, Ceretta L. Health education with adolescents: analysis of knowledge acquisition on health topics. Escola Anna Nery 2015;19(3):484-490 [FREE Full text] [doi: 10.5935/1414-8145.20150064]
7.   European Commission against Racism and Intolerance (ECRI). Section 2: What is Peer Group Education? Definitions, history and methodology. Domino: A manual to use peer group education as a means to fight racism, xenophobia, anti-semitism and intolerance.: Council of Europe URL: https://rm.coe.int/0900001680929871 [accessed 2018-04-19]
8.   Amsellem-Mainguy Y. Qu'entend-on par « éducation pour la santé par les pairs » ? Cahiers de L'action 2014;43:9-16 [FREE Full text]
9.   Peer education and HIV/AIDS: concepts, uses and challenges. UNAIDS Best Practice Collection. Geneva: UNAIDS; 2000 May. URL: https://www.unaids.org/en/resources/documents/2000/20001019_jc291-peereduc_en.pdf [accessed 2020-11-15]
10. Tolli MV. Effectiveness of peer education interventions for HIV prevention, adolescent pregnancy prevention and sexual health promotion for young people: a systematic review of European studies. Health Educ Res 2012 Oct;27(5):904-913. [doi: 10.1093/her/cys055] [Medline: 22641791]
11. Jaworsky D, Larkin J, Sriranganathan G, Clout J, Janssen J, Campbell L, et al. Evaluating Youth Sexual Health Peer Education Programs: Challenges and Suggestions for Effective Evaluation Practices. J Ed Train Stud 2013 Apr;1(1):227-234. [doi: 10.11114/jets.v1i1.68]
12. Sun WH, Miu HYH, Wong CKH, Tucker JD, Wong WCW. Assessing Participation and Effectiveness of the Peer-Led Approach in Youth Sexual Health Education: Systematic Review and Meta-Analysis in More Developed Countries. J Sex Res 2018 Jan 29;55(1):31-44. [doi: 10.1080/00224499.2016.1247779] [Medline: 27898248]
13. Sriranganathan G, Jaworsky D, Larkin J, Flicker S, Campbell L, Flynn S, et al. Peer sexual health education. Health Ed J 2010 Dec 29;71(1):62-71. [doi: 10.1177/0017896910386266]
14. Kim CR, Free C. Recent evaluations of the peer-led approach in adolescent sexual health education: a systematic review. Int Fam Plan Perspect 2008 Jun;34(2):89-96 [FREE Full text] [doi: 10.1363/ifpp.34.0089.08] [Medline: 18644760]
15. Crocker BCS, Pit SW, Hansen V, John-Leader F, Wright ML. A positive approach to adolescent sexual health promotion: a qualitative evaluation of key stakeholder perceptions of the Australian Positive Adolescent Sexual Health (PASH) Conference. BMC Public Health 2019 Jun 03;19(1):681 [FREE Full text] [doi: 10.1186/s12889-019-6993-9] [Medline: 31159767]
16. Wong T, Pharr JR, Bungum T, Coughenour C, Lough NL. Effects of Peer Sexual Health Education on College Campuses: A Systematic Review. Health Promot Pract 2019 Sep;20(5):652-666. [doi: 10.1177/1524839918794632] [Medline: 30141355]
17. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
18. Boyd DM, Ellison N. Social Network Sites: Definition, History, and Scholarship. J Comput Mediated Comm 2007 Oct;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]

19. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman Isler M. Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]

20. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. AIDS Care 2012;24(5):593-600 [FREE Full text] [doi: 10.1080/09540121.2011.630355] [Medline: 22149081]

21. Sun WH, Wong CKH, Wong WCW. A Peer-Led, Social Media-Delivered, Safer Sex Intervention for Chinese College Students: Randomized Controlled Trial. J Med Internet Res 2017 Aug 09;19(8):e284 [FREE Full text] [doi: 10.2196/jmir.7403] [Medline: 28793980]

22. Israel BA, Schulz AJ, Parker EA, Becker AB, Community-Campus Partnerships for Health. Community-based participatory research: policy recommendations for promoting a partnership approach in health research. Educ Health (Abingdon) 2001;14(2):182-197. [doi: 10.1080/13576280110051055] [Medline: 14742017]

23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

24. Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. J Nurs Ed Pract 2016 Jan 15;6(5):11. [doi: 10.5430/jnep.v6n5p100]

25. Lefebvre RC, Flora JA. Social marketing and public health intervention. Health Educ Q 1988 Sep 04;15(3):299-315. [doi: 10.1177/109019818801500305] [Medline: 3056876]

26. Wu D, Tang W, Lu H, Zhang TP, Cao B, Ong JJ, et al. Leading by Example: Web-Based Sexual Health Influencers Among Men Who Have Sex With Men Have Higher HIV and Syphilis Testing Rates in China. J Med Internet Res 2019 Jan 21;21(1):e10171 [FREE Full text] [doi: 10.2196/10171] [Medline: 30664490]

27. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012 Feb 28;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]

28. Patterson SP, Hilton S, Flowers P, McDaid LM. What are the barriers and challenges faced by adolescents when searching for sexual health information on the internet? Implications for policy and practice from a qualitative study. Sex Transm Infect 2019 Sep;95(6):462-467 [FREE Full text] [doi: 10.1136/sextrans-2018-053710] [Medline: 31040251]

29. Brindis CD. A public health success: understanding policy changes related to teen sexual activity and pregnancy. Annu Rev Public Health 2006;27:277-295. [doi: 10.1146/annurev.publhealth.27.021405.102244] [Medline: 16533118]

30. Ralph LJ, Berglas NF, Schwartz SL, Brindis CD. Finding Teens in TheirSpace: Using Social Networking Sites to Connect Youth to Sexual Health Services. Sex Res Soc Policy 2011 Feb 22;8(1):38-49. [doi: 10.1007/s13178-011-0043-4]

31. PASH: Positive Adolescent Sexual Health. URL: http://www.pash.org.au/ [accessed 2020-11-10]

32. Syred J, Naidoo C, Woodhall SC, Baraitser P. Would you tell everyone this? Facebook conversations as health promotion interventions. J Med Internet Res 2014 Apr 11;16(4):e108 [FREE Full text] [doi: 10.2196/jmir.3231] [Medline: 24727742]

33. Wise K, Hamman B, Thorson K. Moderation, Response Rate, and Message Interactivity: Features of Online Communities and Their Effects on Intent to Participate. J Comp Mediated Comm 2006 Oct;12(1):24-41. [doi: 10.1111/j.1083-6101.2006.00313.x]

34. Milburn K. A critical review of peer education with young people with special reference to sexual health. Health Educ Res 1995 Dec;10(4):407-420. [doi: 10.1093/her/10.4.407] [Medline: 10159674]

35. Simoni JM, Franks JC, Lehavot K, Yard SS. Peer interventions to promote health: conceptual considerations. Am J Orthopsychiatry 2011 Jul;81(3):351-359 [FREE Full text] [doi: 10.1111/j.1939-0025.2011.01103.x] [Medline: 21729015]

36. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]

37. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]

38. Campo S, Askelson NM, Spies EL, Losch M. Caution, The Use of Humor May Lead to Confusion: Evaluation of a Video Podcast of the Midwest Teen Sex Show. Am J Sex Ed 2010 Aug 31;5(3):201-216. [doi: 10.1080/15546128.2010.503857]

39. Nicholas A, Bailey JV, Stevenson F, Murray E. The Sexunzipped trial: young people's views of participating in an online randomized controlled trial. J Med Internet Res 2013 Dec 12;15(12):e276 [FREE Full text] [doi: 10.2196/jmir.2647] [Medline: 24334198]

40. Brady SS, Sieving RE, Terveen LG, Rosser BRS, Kodet AJ, Rothberg VD. An Interactive Website to Reduce Sexual Risk Behavior: Process Evaluation of TeensTalkHealth. JMIR Res Protoc 2015 Sep 02;4(3):e106 [FREE Full text] [doi: 10.2196/resprot.3440] [Medline: 26336157]

41. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: A Social Media-Based Intervention Designed to Increase HIV Care Linkage, Retention, and Health Outcomes for Racially and Ethnically Diverse Young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]

XSL•FO
RenderX

42.   Platt LF, Lenzen AL. Sexual orientation microaggressions and the experience of sexual minorities. J Homosex 2013;60(7):1011-1034. [doi: 10.1080/00918369.2013.774878] [Medline: 23808348]

43.   Carrotte ER, Vella AM, Bowring AL, Douglass C, Hellard ME, Lim MSC. "I am yet to encounter any survey that actually reflects my life": a qualitative study of inclusivity in sexual health research. BMC Med Res Methodol 2016 Jul 27;16:86 [FREE Full text] [doi: 10.1186/s12874-016-0193-4] [Medline: 27465507]

Abbreviations

INSERM: Institut National de la Santé Et de la Recherche Médicale
LGBT: lesbian, gay, bisexual, transsexual
MSM: men who have sex with men

Original Paper

Computer-Tailored Decision Support Tool for Lung Cancer Screening: Community-Based Pilot Randomized Controlled Trial

Lisa Carter-Harris[1*], PhD; Robert Skipworth Comer[2*], MSc; James E Slaven II[3*], MSc; Patrick O Monahan[3*], PhD; Emilee Vode[4*], BSc; Nasser H Hanna[3*], MD; DuyKhanh Pham Ceppa[3*], MD; Susan M Rawl[4*], PhD

[1]Memorial Sloan Kettering Cancer Center, New York, NY, United States

[2]Indiana University School of Informatics and Computing, Indianapolis, IN, United States

[3]School of Medicine, Indiana University, Indianapolis, IN, United States

[4]Indiana University School of Nursing, Indianapolis, IN, United States

[*]all authors contributed equally

Corresponding Author:
Lisa Carter-Harris, PhD
Memorial Sloan Kettering Cancer Center
641 Lexington Avenue
7th Floor
New York, NY, 10022
United States
Phone: 1 6468880040
Email: carterl@mskcc.org

Abstract

Background: Lung cancer screening is a US Preventive Services Task Force Grade B recommendation that has been shown to decrease lung cancer-related mortality by approximately 20%. However, making the decision to screen, or not, for lung cancer is a complex decision because there are potential risks (eg, false positive results, overdiagnosis). Shared decision making was incorporated into the lung cancer screening guideline and, for the first time, is a requirement for reimbursement of a cancer screening test from Medicare. Awareness of lung cancer screening remains low in both the general and screening-eligible populations. When a screening-eligible person visits their clinician never having heard about lung cancer screening, engaging in shared decision making to arrive at an informed decision can be a challenge. Methods to effectively prepare patients for these clinical encounters and support both patients and clinicians to engage in these important discussions are needed.

Objective: The aim of the study was to estimate the effects of a computer-tailored decision support tool that meets the certification criteria of the International Patient Decision Aid Standards that will prepare individuals and support shared decision making in lung cancer screening decisions.

Methods: A pilot randomized controlled trial with a community-based sample of 60 screening-eligible participants who have never been screened for lung cancer was conducted. Approximately half of the participants (n=31) were randomized to view LungTalk—a web-based tailored computer program—while the other half (n=29) viewed generic information about lung cancer screening from the American Cancer Society. The outcomes that were compared included lung cancer and screening knowledge, lung cancer screening health beliefs (perceived risk, perceived benefits, perceived barriers, and self-efficacy), and perception of being prepared to engage in a discussion about lung cancer screening with their clinician.

Results: Knowledge scores increased significantly for both groups with greater improvement noted in the group receiving LungTalk (2.33 vs 1.14 mean change). Perceived self-efficacy and perceived benefits improved in the theoretically expected directions.

Conclusions: LungTalk goes beyond other decision tools by addressing lung health broadly, in the context of performing a low-dose computed tomography of the chest that has the potential to uncover other conditions of concern beyond lung cancer, to more comprehensively educate the individual, and extends the work of nontailored decision aids in the field by introducing tailoring algorithms and message framing based upon smoking status in order to determine what components of the intervention drive behavior change when an individual is informed and makes the decision whether to be screened or not to be screened for lung cancer.

International Registered Report Identifier (IRRID):   RR2-10.2196/resprot.8694

XSL•FO

RenderX

Introduction

Lung cancer screening is recommended by the US Preventive Services Task Force (USPSTF) with a Grade B recommendation and offers the potential to detect lung cancer via low-dose computed tomography of the chest at an earlier stage when more treatment options exist [1]. However, lung cancer screening is complex because there are associated risks and potential harms that complicate the decision to screen [1]. False-positive results and incidental findings have the potential to lead to invasive testing that are possible in this type of cancer screening [1,2]. Because of these potential risks, lung cancer screening is a preference-sensitive decision and requires patient-clinician discussion and a shared decision-making process. Therefore, the USPSTF recommends, and Medicare requires, that the decision to screen for lung cancer be the result of a documented shared decision making and counseling visit between a patient and their clinician with the use of one or more decision aids [1,3].

Knowledge and awareness about lung cancer screening among the general population is extremely low [4,5]. In order to foster increased patient-clinician discussions about this relatively new screening option, it is essential to leverage new ways to both increase awareness and knowledge about lung cancer screening. In response to our prior work with screening-eligible individuals, our team developed a theoretically grounded, computer-tailored decision support tool titled LungTalk (1) to increase awareness about lung cancer screening and improve knowledge about lung health and lung cancer screening benefits and risks, and (2) prepare individuals to engage in discussions with their clinician in order to support the shared decision-making process.

In this paper, the results of a community-based, pilot randomized controlled trial to compare the effects of LungTalk to those of a nontailored lung screening information sheet in a sample of screening-eligible individuals are presented. LungTalk was developed using the USPSTF Lung Cancer Screening Guidelines as well as qualifying and certification criteria of the International Patient Decision Aid Standards instrument [6] as described in our protocol [7]. In addition to evaluating feasibility of the study procedures, the following research questions were answered:

1. Are there changes in knowledge of lung cancer risk and screening, and lung cancer screening health beliefs (perceived risk, perceived benefits, perceived barriers, self-efficacy) between patients who received LungTalk and those who received the nontailored lung screening information sheet?
2. Are there changes in participants' perceptions of being prepared to engage in a discussion with their clinician about lung cancer screening between patients who received LungTalk and those who received the nontailored lung screening information sheet?
3. Are there changes in self-reported patient-clinician discussions about lung cancer screening and receipt of a lung cancer screening recommendation between patients who received LungTalk and those who received the nontailored lung screening information sheet?

Methods

Study Sample and Recruitment

Participants (n=60), both men and women, who were eligible for lung cancer screening were recruited using Facebook-targeted advertisement. (Facebook has the ability to "target" an advertisement by demographics and keywords listed in each individual Facebook user's profile or interest list.) Using this technique, we were able to purposively sample people aged 55 years and older who indicated smoking as an interest or like in their profile. Frandsen et al [8] demonstrated that participants recruited to smoking cessation randomized clinical trials with Facebook advertisements did not differ from those recruited by traditional methods in either smoking characteristics or demographics. Our team extended Fransden's [8] work by demonstrating that participants aged 55 years and older recruited to a web-based survey study for lung cancer screening did not differ from those recruited by traditional methods by either smoking characteristics or demographics thus demonstrating its utility to reach and recruit older long-term smokers [9]. A REDCap (Research Electronic Data Capture; Vanderbilt University) survey was used to screen and invite eligible participants to enroll in the study. Inclusion criteria mirrored the USPSTF Lung Cancer Screening Guidelines [1]: (1) aged 55 to 80 years; (2) 30 pack-year tobacco smoking history; (3) current smoker or former smoker who quit within the past 15 years; (4) not diagnosed with a condition that would be contraindicated for lung cancer screening; and (5) not diagnosed with lung cancer.

Interventions

LungTalk Interactive Program

LungTalk is a computer-tailored decision support tool that is theoretically grounded in the Conceptual Model on Lung Cancer Screening Participation [10]. This model links the Health Belief Model [11] to the Precaution Adoption Process Model [12] and includes key psychological variables (eg, stigma, mistrust, fatalism, fear, and worry) as factors that may influence an individual's decision to screen, or not, for lung cancer [10]. LungTalk is designed to increase knowledge and awareness about the option to screen, or not, for lung cancer and to prepare screening-eligible individuals to engage in shared decision making about lung cancer screening with their clinician. LungTalk educates high-risk individuals about (1) lung health broadly including the effects of nicotine; (2) risk factors for the development of lung cancer; (3) the option of lung cancer screening with low-dose computed tomography of the chest; and (4) the risks and benefits of lung cancer screening.

Furthermore, because our prior research [4,13] revealed that messages needed to be different for needed to be different for individuals who currently versus used to smoke, messages in LungTalk are tailored by smoking status (see Multimedia Appendix 1).

LungTalk includes audio, video, and animation segments with tailoring algorithms for scripts presented from a master content library. In addition, LungTalk offers the option of saving or printing a tailored summary (at program completion) that individuals can use to guide a discussion with their clinician. This summary highlights key points related to lung health and screening tailored by smoking status, offers questions the user can ask to initiate a discussion with their clinician, and includes specific questions identified by the user that they wish to discuss with their clinician. Content in LungTalk is visually presented with text written at an eighth grade reading level. To meet the needs of people with low literacy and auditory preference learning styles, all content is narrated as well as shown as written text on screens (see Multimedia Appendix 1). Additional details on the development of LungTalk are described elsewhere [7].

Nontailored Lung Screening Information Sheet

The control group viewed a nontailored lung screening information sheet online that contained information compiled from lung cancer screening information developed by the American Cancer Society. The reading level of this written material was at an eighth grade level.

Data Collection

Data collected between January 2017 and February 2017 from participants in the state of Indiana. Data were collected via online surveys completed by participants (baseline only) and telephone interviews conducted by trained research staff. The follow-up surveys were developed in REDCap, a secure web-based application to build and manage online surveys and databases. Participants completed a 20-minute baseline survey prior to randomization. Follow-up surveys were then completed 1 week and 3 months postintervention.

The baseline survey collected data on sociodemographic and health status characteristics, lung cancer and screening knowledge, lung cancer screening health beliefs (perceived risk of lung cancer, perceived benefits of, perceived barriers to, and self-efficacy for lung cancer screening) [11], perceived preparation to engage in a patient-clinician discussion about lung cancer screening, and stage of adoption for lung cancer screening. At completion of the baseline survey, each participant was randomized to receive either LungTalk or the nontailored lung screening information sheet. A link to either LungTalk or the lung screening information sheet was emailed to the participant based upon their randomization.

A follow-up telephone interview was conducted within 1 week of the participant completing the intervention. The interview included items to assess lung cancer and screening knowledge, lung cancer screening health beliefs (perceived risk, perceived benefits, perceived barriers, and self-efficacy) [11], satisfaction with the intervention, self-report of perception of preparation to engage in a patient-clinician discussion about lung cancer screening, and stage of adoption for lung cancer screening [12].

A second telephone interview was completed 3 months after receipt of the intervention to assess whether the participant had a subsequent discussion about lung cancer screening with a clinician, received a clinician recommendation for screening, and stage of adoption for lung cancer screening [12].

Measures

Guided by the Conceptual Model on Lung Cancer Screening Participation, valid and reliable instruments were used to measure knowledge of lung cancer risk and screening, lung cancer screening health beliefs (perceived risk of lung cancer, perceived benefits of, perceived barriers to, and self-efficacy for lung cancer screening), health care clinician recommendation, perception of preparation to engage in a patient-clinician discussion about lung cancer screening, and stage of adoption for lung cancer screening participation. Stage of adoption was measured using an algorithm that is theoretically based upon the Precaution Adoption Process model [12]. The Precaution Adoption Process Model is supported by the Conceptual Model on Lung Cancer Screening Participation as an appropriate outcome variable in lung cancer screening decisions because it includes a stage to categorize individuals who have thoroughly weighed their options and decided not to be screened [12]. The Precaution Adoption Process Model categorizes individuals into 1 of 7 stages, including unaware, aware but unengaged, undecided, decided not to act, decided to act, action, and maintenance [14]. The Lung Cancer Screening Health Belief Scales (perceived risk, perceived benefits, perceived barriers, self-efficacy) had been previously validated in a community-based sample of 497 screening-eligible individuals [11]. *Total knowledge*, with Knowledge of Lung Cancer and Lung Cancer Screening, which is a 6-item multidimensional scale used in our preliminary study adapted from literature specific to lung cancer, will be assessed, including knowledge of lung cancer, risk, and screening; *total perceived risk*, with Perceived Risk of Lung Cancer Scale which is a 3-item scale with higher scores indicative of higher perceived risk of lung cancer; *total perceived benefits* (Cronbach α=.90), with Perceived Benefits of Lung Cancer Screening Scale which is a 6-item scale with higher scores reflective of higher perceived benefits of lung cancer screening (Cronbach α=.68); *total perceived barriers*, with Perceived Barriers to Lung Cancer Screening Scale which is a 17-item scale where higher scores reflect higher perceived barriers to lung cancer screening (Cronbach α=.86); and *total self-efficacy*, with Self-Efficacy for Lung Cancer Screening Scale and is a 9-item scale to assess individual beliefs about ability to arrange and complete an low-dose computed tomography to screen for lung cancer with higher scores reflective of higher levels of self-efficacy for lung cancer screening (Cronbach α=.83).

Statistical Analysis

Deidentified data collected via REDCap were exported for analyses. Data completeness was assessed through descriptive analyses. Means and standard deviations or frequency distributions were examined to check for coding errors and out-of-range values.

Our first goal was to evaluate the feasibility of study procedures. Therefore, we calculated study participation rates and rates of

completion and retention of participants at baseline (T1), 1-week postintervention (T2), and 3 months postintervention (T3). For each, we calculated the proportion of people who were recruited initially and retained at each data collection timepoint (see Figure 1). Patterns of missing values were examined and evaluated for randomness as described by Enders [15]. Diagnostic plots and inferential tests for tenability of

assumptions were evaluated, and appropriate remedial methods were applied where required. We calculated means and standard deviations for each of the key study variables at each time point, as well as change scores calculated by value at T2 minus the value at T1, for both intervention groups. Means and standard deviations of study variables by group were also calculated.

Figure 1. Participant recruitment flowchart.
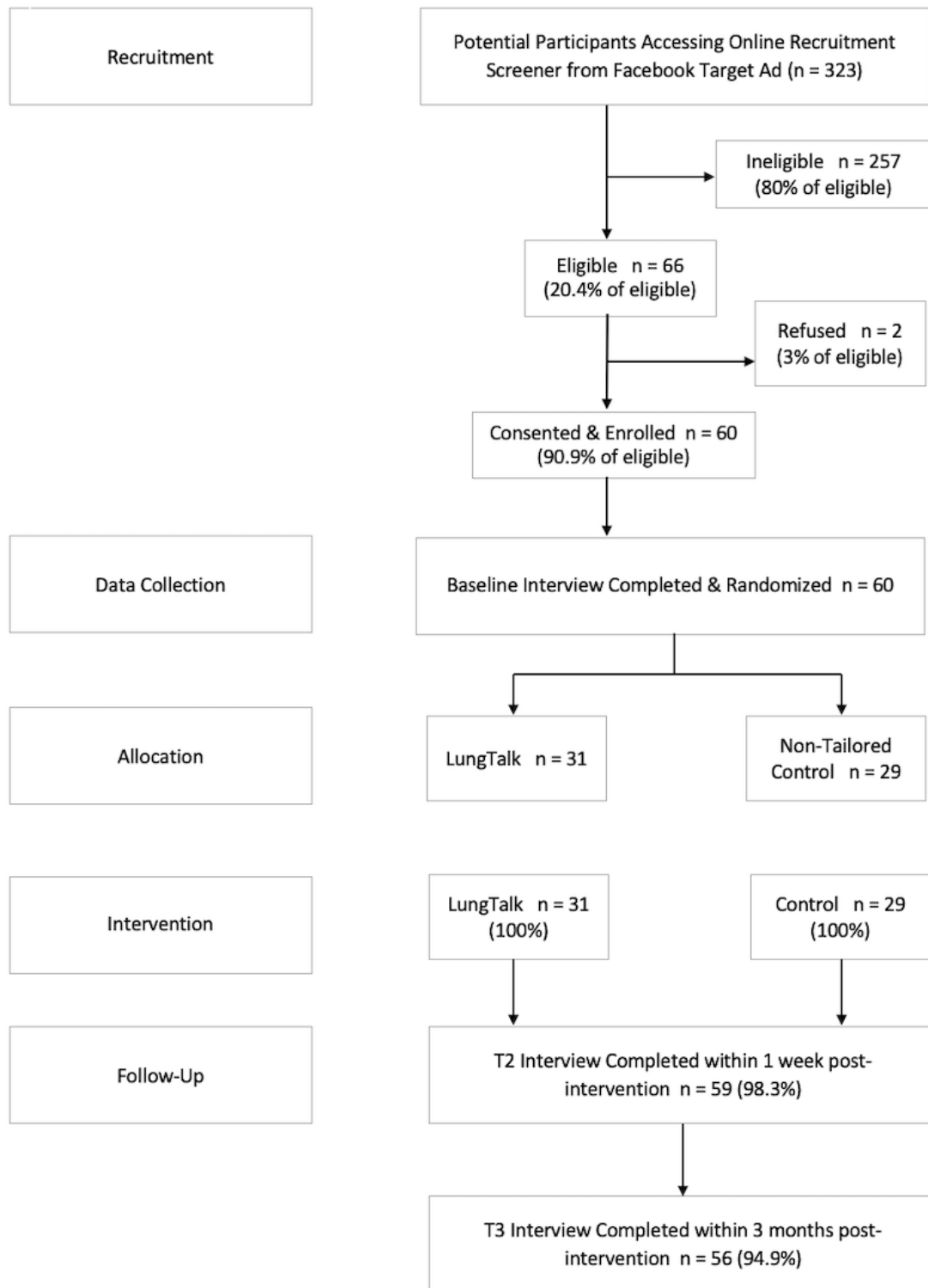
Analysis of covariance models (ANCOVAs) were performed to determine if there were significant differences in change

scores between groups, adjusted for the value at T1. Using the model result's means and standard error, we calculated the

pooled model-based standard deviation and Cohen *d* effect sizes (ie, difference between groups on the mean change scores of study variables divided by the model-based pooled SD of change scores). All analytic assumptions were verified, and analyses were performed using SAS (version 9.4; SAS Institute). Of the 60 participants recruited, 31 were randomized to receive LungTalk and 29 were randomized to the nontailored lung screening information sheet.

Results

Participant Sociodemographics

Participants ranged in age from 55 to 74 years (mean 62.4, SD 5.2) with fairly equal numbers of men (29/60, 48%) and women (31/60, 52%). The majority were White (48/60, 80%). Participant sociodemographic and health status characteristics are shown by intervention group in Table 1. No significant differences in sociodemographic or smoking status characteristics were observed. To assess feasibility of study procedures, Figure 1 shows the flow of participants through the study. We interviewed 100% (31/31) of group 1 (LungTalk) and 98.3% (28/29) of group 2 (nontailored lung screening information sheet) within 1 week of intervention completion.

Table 1. Demographics and smoking characteristics at baseline.

Characteristic	Overall (n=60)	LungTalk (n=31)	Nontailored lung screening information sheet (n=29)	P value
Age (years), mean (SD)	62.2 (5.2)	61.2 (4.8)	63.2 (5.5)	.13[a]
Age (years), n (%)				.07[b]
55-64	38 (63)	23 (74)	15 (52)	
65+	22 (37)	8 (26)	14 (48)	
Sex, n (%)				.99[b]
Male	29 (48)	15 (48)	14 (48)	
Female	31 (52)	16 (52)	15 (52)	
Race, n (%)				.86[c]
White	48 (80)	24 (77)	24 (83)	
Black	10 (17)	6 (19)	4 (14)	
Other	2 (3)	1 (3)	1 (3)	
Family history of lung cancer, n (%)				.73[c]
Yes	10 (17)	6 (19)	4 (14)	
No	50 (83)	25 (81)	25 (86)	
Smoking status, n (%)				.78[b]
Former	28 (46.7)	15 (48.4)	13 (44.8)	
Current	32 (53.3)	16 (51.6)	16 (55.2)	
Education, n (%)				.23[c]
<High school	1 (2)	1 (3)	0 (0)	
High school/GED	19 (32)	10 (32)	9 (31)	
Some college	19 (32)	12 (39)	7 (24)	
College graduate	21 (35)	8 (26)	13 (45)	
Income, n (%)				
<$25,000	25 (42)	14 (45)	11 (39)	.94[b]
$25,000-$50,000	20 (34)	9 (29)	11 (39)	
>$50,000	14 (24)	8 (26)	6 (21)	
Health insurance, n (%)				.33[c]
Medicare	14 (24)	9 (30)	5 (18)	
Medicaid	8 (14)	4 (13)	4 (14)	
Private	18 (31)	11 (37)	7 (25)	
Medicare+supplement	8 (14)	2 (7)	6 (21)	
Other	2 (3.5)	0 (0)	2 (7)	
Multiple	8 (14)	4 (13)	4 (14)	
Pack-years tobacco smoking, mean (SD)	48.7 (19.5)	47.6 (21.9)	49.9 (16.6)	.29[a]
Packs smoked daily, mean (SD)	1.3 (0.5)	1.4 (0.5)	1.3 (0.5)	.97[a]
Years smoked, mean (SD)	36.6 (8.3)	35.4 (9.3)	37.9 (7.0)	.10[a]
Years since quitting, mean (SD)	6.3 (5.0)	6.7 (5.7)	5.8 (4.4)	.77[a]

[a]Wilcoxon rank-sum test.

[b]Chi-square test.

XSL•FO

RenderX

[c]Fisher exact test.

Changes in Knowledge

Mean scores, standard deviations, and change scores for knowledge and beliefs are shown in Table 2, along with the effect sizes and tests of the within-group change. Knowledge scores increased significantly for both groups from baseline to postintervention, although the improvement was greater for the group receiving LungTalk (2.33 vs 1.14 change in means which represents a 1.5 SD vs 1 SD change). Perceived self-efficacy for lung cancer screening also increased significantly for both groups, with a 0.5 SD change for each group. A small reduction in perceived risk of getting lung cancer was observed in the LungTalk group (SRM –0.13 or about one-tenth of an SD decrease) while an increase in perceived risk was seen in the nontailored lung screening information sheet group (SRM 0.30 or about one-third of an SD increase). Perceived benefits improved significantly for the LungTalk group (SRM 0.41, almost one-half SD increase), whereas the minor decrease in benefits was not significant for the nontailored lung screening information sheet group (SRM –0.14). A nonsignificant reduction in perceived barriers was observed in both groups, although the reduction was greater for the enhanced control group (SRM –0.33 vs –0.15).

While Table 2 provides insight into the within-group effects from each group separately, estimates of efficacy of LungTalk on knowledge and health beliefs are presented in Table 3. In Table 3, the between-group effect sizes and P values are shown for comparing groups on changes from baseline to 1-week postintervention on total scale scores, controlling for T1 value of the score (ie, controlling for initial score level at baseline). Compared with the nontailored lung screening information sheet group, the LungTalk group had a greater increase in total knowledge scale scores (8 items with P<.01). The large effect

size of 0.85 indicated that the increase in knowledge for those who received LungTalk was more than three-quarters of an SD greater than the increase in knowledge for those who received the nontailored lung screening information sheet. While the change in perceived benefits did not reach significance, change was occurring in the theoretically expected direction. LungTalk increased participants' perceptions of the benefits of lung cancer screening while benefits scores decreased slightly for the lung screening information sheet group (1.07 vs –0.18, P=.06).

Because knowledge was significantly higher for those who received LungTalk, we examined group differences on individual knowledge items to see where specific improvements were made (Table 4). Interestingly, the percentage of participants who correctly answered that a low-dose computed tomography of the chest is the test that is currently recommended for lung cancer screening increased significantly from baseline in both groups (LungTalk: 57% increase, lung screening information sheet: 52% increase; P<.001). Likewise, the percentages of participants in both groups who correctly knew that lung cancer screening is only recommended for current and former smokers increased significantly from baseline (LungTalk: 38%, P=.004; lung screening information sheet: 24%; P=.008). Compared to the lung screening information sheet group, significantly greater improvements in knowledge were seen for those receiving LungTalk on 3 knowledge items: (1) knowing that a person should talk with their health care provider about lung cancer screening before being screened (P=.10 vs P=.01, respectively); (2) knowing that lung cancer screening, if results are normal, should be done annually (P=.71 vs P=.007, respectively); and (3) knowing that 55 is the age that people should start screening for lung cancer (P=.48 vs P<.001, respectively).

Table 2. Scores at baseline (T1) and 1-week postintervention (T2) and within-group tests.

Variable	Lung Talk, mean (SD)	Nontailored lung screening information sheet, mean (SD)
Total knowledge score		
T1	3.90 (1.47)	3.66 (1.47)
T2	6.27 (1.26)	4.79 (1.32)
Change	2.33 (1.54)	1.14 (1.16)
SRM[a]	1.51	0.98
P value[b]	<.01	<.01
Total perceived risk		
T1	13.74 (2.66)	13.69 (1.95)
T2	13.43 (2.40)	14.28 (2.36)
Change	–0.27 (2.13)	0.59 (1.96)
SRM	–0.13	0.30
P value[b]	.50	.12
Total perceived benefits		
T1	17.55 (1.88)	18.34 (2.70)
T2	18.70 (3.10)	18.07 (2.89)
Change	1.17 (2.85)	–0.28 (2.00)
SRM	0.41	–0.14
P value[b]	.03	.46
Total perceived barriers		
T1	34.10 (7.15)	33.03 (6.48)
T2	33.20 (7.43)	30.90 (7.24)
Change	–0.90 (5.86)	–2.14 (6.49)
SRM	–0.15	–0.33
P value[b]	.41	.09
Total self-efficacy		
T1	27.45 (5.08)	28.38 (4.87)
T2	28.97 (4.55)	30.38 (4.16)
Change	1.53 (3.30)	2.00 (3.56)
SRM	0.46	0.56
P value[b]	.02	.01

[a]SRM: standardized response mean = mean change / SD of change.

[b]2-sided paired test.

Table 3. Effect sizes.

Variable	Scale range	Adjusted mean (SE) change		Comparison			
		LungTalk	Nontailored lung screening information sheet	F test ($df1$, $df2$)	P value	Mean difference (95% CI)	Cohen d effect size[a]
Total knowledge	0-6	2.41 (0.20)	1.06 (0.21)	21.5 (1, 56)	<.01	1.35 (0.77, 1.93)	0.8482
Total perceived risk	3-12	–0.26 (0.34)	0.58 (0.35)	3.03 (1, 56)	.09	–0.85 (–1.83, 0.13)	–0.3179
Total perceived benefits	6-24	1.07 (0.45)	–0.18 (0.45)	3.79 (1, 56)	.06	1.25 (–0.04, 2.54)	0.3555
Total perceived barriers	17-68	–0.72 (1.06)	–2.32 (1.07)	1.12 (1, 56)	.29	1.60 (–1.43, 4.62)	0.1931
Total self-efficacy	9-36	1.37 (0.54)	2.17 (0.55)	1.08 (1, 56)	.30	–0.80 (–2.35, 0.75)	–0.1896

[a]Positive effect size indicates greater increase from T1 to T2 for LungTalk than for InfoSheet. Negative effect size indicates greater increase from T1 to T2 for InfoSheet than for LungTalk.

Table 4. Individual knowledge items by within-group change.

Knowledge item	LungTalk			Nontailored lung screening information sheet		
	Baseline, n (%) correct	T2, n (%) correct	P value[a]	Baseline, n (%) correct	T2, n (%) correct	P value[a]
Who is more likely to get lung cancer? (a person who has smoked cigarettes for a long time)	27 (87.1)	28 (93.3)	.41	26 (89.7)	27 (93.1)	.56
What is the most common symptom of lung cancer? (chronic cough)	25 (80.7)	25 (83.3)	.56	24 (82.8)	27 (93.1)	.18
Which test is currently recommended for lung cancer screening? (low-dose CT[b] scan)	8 (25.8)	25 (83.3)	<.001	9 (31.0)	24 (82.8)	.001
Compared to a chest x-ray, how much radiation does a lung scan expose you to? (about the same as a chest x-ray)	6 (19.4)	11 (36.7)	.06	6 (20.7)	8 (27.6)	.41
What should a person do before being screened for lung cancer? (talk with their health care provider about low-dose CT screening)	21 (67.7)	29 (96.7)	.01	20 (69.0)	24 (82.8)	.10
If you choose to have a lung scan to screen for lung cancer and everything is normal, when will you need to have your next one? (in 1 year)	17 (54.8)	25 (83.3)	.007	8 (27.6)	7 (24.1)	.71
Who is currently recommended to have a lung scan to screen for lung cancer? (only current and former smokers)	10 (32.3)	21 (70.0)	.004	6 (20.7)	13 (44.8)	.008
At what age is it recommended that people start to screen for lung cancer? (55)	7 (22.6)	24 (80.0)	<.001	7 (24.1)	9 (31.0)	.48

[a]2-sided McNemar test of paired proportions.

[b]CT: computed tomography.

Changes in Participants' Perceptions

As shown in Table 5, satisfaction with the LungTalk intervention was significantly higher than with the nontailored lung sreening information sheet. Individuals in both groups felt "prepared" or "very prepared" to have a discussion with their clinician about lung screening, with no significant differences between the 2 intervention groups on preparedness (P=.52).

Table 5. User satisfaction, clinician recommendation, shared decision-making discussion by group.

Variable	Lung Talk (n=31)	Nontailored lung screening information sheet (n=29)	P value[a]
Satisfaction (T2)			.002
Not at all satisfied	0 (0)	1 (3)	
Somewhat satisfied	1 (3)	4 (14)	
Satisfied	7 (23)	16 (55)	
Very satisfied	22 (73)	8 (28)	
Preparedness (T2)			.52
Somewhat prepared	4 (13)	6 (21)	
Prepared	10 (33)	6 (21)	
Very prepared	16 (53)	17 (59)	
Clinician recommendation (T3)			.33
Yes	8 (28)	4 (15)	
No	22 (72)	25 (85)	
Shared decision-making discussion about lung cancer screening (T3)			.23
Yes	10 (34)	5 (19)	
No	20 (66)	24 (81)	

[a]2-sided Fisher exact test.

P-value is from two-sided Fisher's exact Test.

Changes in Self-Reported Patient-Clinician Discussions

At 6 months, though the number of participants who reported that they had a discussion with their clinician in the LungTalk group (10/31, 34.5%) was double that in the nontailored lung screening information sheet group (5/29, 18.5%), this difference was not significant (*P*=.23). Similar results were observed at 6 months for receipt of a clinician recommendation for lung screening; 27.6% (8/31) in the LungTalk group reported receiving a recommendation from their clinician compared to 14.8% (4/29) in the nontailored lung screening information sheet group. This difference was also not significant (*P*=.33).

Discussion

General

It is still relatively early on in the development of decision support tools for lung cancer screening. Most have focused on calculating personal risk for the development of lung cancer and subsequent recommendations to screen are based upon calculated risk status [16-20]. These tools range in level of complexity and delivery including pamphlets, brochures, videos, educational scripts, and computer programs [16-20]. These tools can also be deployed in multiple formats such as by mail, telephone, in person, and via the internet. Dharod and colleagues [14] examined the feasibility of a digital health outreach strategy via a patient portal directing individuals to an interactive website which then accessed screening eligibility.

Similar to other patient decision aids [16-20], the mPATH (mobile Patient Technology for Health) Lung Interactive website is atheoretical and calculates risk for lung cancer but does not tailor beyond personalized risk [14]. LungTalk focuses on empowering individual patients with knowledge so that they are an informed partner as they discuss with their clinician and make a decision about lung cancer screening. In addition, LungTalk goes beyond much of what other decision tools focus on by addressing lung health broadly, in the context of performing a low-dose computed tomography of the chest that has the potential to uncover other conditions of concern beyond lung cancer, to more comprehensively educate the individual. LungTalk extends the work of nontailored decision aids in the field, by introducing tailoring algorithms and message framing based upon smoking status in order to determine what components of the intervention are driving behavior change when an individual is informed and makes the decision to screen, or not, for lung cancer. LungTalk is innovative in its tailored messaging approach based on smoking status and tailored printout that helps patients initiate a discussion with their clinician about their lung health and the option of screening.

As an intervention, individuals using LungTalk felt equally prepared to engage in a discussion with their clinician about lung cancer screening as they did with the nontailored lung screening information sheet. From the patient perspective, being prepared to engage in a shared discussion about lung cancer screening is essential to successfully involving the patient in the dyadic communication clinical context of this patient-clinician equitable engagement; however, participants were significantly more satisfied with LungTalk (*P*=.002). As interventions are developed, it is essential for developers to take into consideration user satisfaction in efforts to increase the likelihood of both initial and sustained engagement with the intervention.

Consistent with other types of cancer screening, knowledge and health beliefs have been shown to be associated with lung cancer screening behavior [11]. In addition, there is strong support in

other types of cancer screening that tailored interventions are much more effective in promoting cancer screening behavior because messaging is more personally relevant to the individual making the decision to screen. LungTalk included messaging to increase knowledge and results support its effectiveness in doing so over the nontailored intervention (8 item with $P<.01$). In particular, knowing that normal or negative lung cancer screening results still require adherence to annual screening while eligible is important and LungTalk increased this knowledge level compared to the change from the nontailored lung screening information sheet ($P=.007$). Since the majority of lung cancers identified on lung cancer screening exams were on subsequent exams as opposed to initial exams, this is essential knowledge for screening-eligible individuals.

With regard to health beliefs, even though LungTalk improved perceived benefits and self-efficacy in expected directions, changes were not significantly different compared those of the lung screening information sheet. LungTalk fell short in reducing perceived barriers to lung cancer screening, an important variable that often predicts cancer screening behaviors. The impact of these interventions on perceived risk of lung cancer was not observed and highlights a critical gap in the tailored messaging component of the intervention. Moving forward, it is important that tailored messages be further refined to target specific barriers to lung cancer screening as well as perceived risk as it relates to lung cancer for the target patient population in order to improve the efficacy of LungTalk as both a health communication and decision support tool.

Strengths and Limitations

As a pilot study, we had adequate power for estimating effect sizes and detecting large effect sizes. For example, 26 participants in each group were required for 80% power to detect a large Cohen d effect size of 0.80 SD difference between means, and we slightly exceeded 26 per group. A larger sample (eg, 64 per group) would be needed to detect a medium effect size of 0.50 with 80% power.

As with all studies, this study is not without limitations. Our recruitment methods may have influenced who participated in this study. Targeted Facebook advertisement allowed us to purposively sample people aged 55 years and older who indicated smoking as an interest on their Facebook page. People who use Facebook and who indicate smoking as an interest may constitute unique sample. Our sample demographics, however, indicate that we successfully recruited a racially diverse, national sample with equal numbers of men and women. Randomization also was effective; no differences between groups were observed at baseline. Our nonsignificant group differences in change scores on health belief variables were likely due to inadequate integration of content and tailored messages in LungTalk that would impact health belief constructs. Future versions of LungTalk need to address these constructs specifically in content and tailored messages if we are to change health beliefs in the directions that promote patient-clinician discussions about lung cancer screening and shared decisions.

Conclusion

Preliminary results indicate LungTalk is a helpful communication tool for individuals who are considering the option of lung cancer screening. Specifically, LungTalk can help enhance the shared decision-making process by priming individuals with essential baseline knowledge to support an informed discussion with a health care clinician about potential risks and benefits related to lung cancer screening.

Authors' Contributions

LCH and SMR conceived of the study and sought funding and ethical approval. LCH and SMR are responsible for the development phase with the Community Advisory Board. LCH, EV, and RSC are responsible for the usability testing phase of the study. LCH and SMR are responsible for the pilot randomized controlled trial phase and planned statistical analysis. LCH, EV, SMR, and RSC assisted in the development of LungTalk and subsequent iterations. NH and DPC provided feedback on the concept of LungTalk at key points in development. EV has assisted with data collection in all phases of the testing. All authors have been involved in drafting and revising the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots - LungTalk.
[PDF File (Adobe PDF File), 3264 KB - jmir_v22i11e17050_app1.pdf]

References

1.   Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014 Mar 4;160(5):330-338. [doi: 10.7326/M13-2771] [Medline: 24378917]

2.   National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 4;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]

3.   Decision memo for screening for lung cancer with low-dose computed tomography (LDCT). 2015. URL: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274 [accessed 2020-02-15]

4.   Carter-Harris L, Ceppa DP, Hanna N, Rawl SM. Lung cancer screening: what do long-term smokers know and believe? Health Expect 2017 Feb;20(1):59-68 [FREE Full text] [doi: 10.1111/hex.12433] [Medline: 26701339]

5.   Carter-Harris L, Tan ASL, Salloum RG, Young-Wolff KC. Patient-provider discussions about lung cancer screening pre- and post-guidelines: Health Information National Trends Survey (HINTS). Patient Educ Couns 2016 Nov;99(11):1772-1777. [doi: 10.1016/j.pec.2016.05.014] [Medline: 27241830]

6.   Joseph-Williams N, Newcombe R, Politi M, Durand M, Sivell S, Stacey D, et al. Toward minimum standards for certifying patient decision aids: a modified delphi consensus process. Med Decis Making 2014 Aug;34(6):699-710. [doi: 10.1177/0272989X13501721] [Medline: 23963501]

7.   Carter-Harris L, Comer RS, Goyal A, Vode EC, Hanna N, Ceppa D, et al. Development and usability testing of a computer-tailored decision support tool for lung cancer screening: study protocol. JMIR Res Protoc 2017 Nov 16;6(11):e225 [FREE Full text] [doi: 10.2196/resprot.8694] [Medline: 29146565]

8.   Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]

9.   Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]

10.   Carter-Harris L, Davis LL, Rawl SM. Lung cancer screening participation: developing a conceptual model to guide research. Res Theory Nurs Pract 2016 Nov 01;30(4):333-352. [doi: 10.1891/1541-6577.30.4.333] [Medline: 28304262]

11.   Carter-Harris L, Slaven JE, Monohan P, Rawl SM. Development and psychometric evaluation of the Lung Cancer Screening Health Belief Scales. Cancer Nurs 2017;40(3):237-244 [FREE Full text] [doi: 10.1097/NCC.0000000000000386] [Medline: 27244666]

12.   Weinstein N, Sandman P, Blalock S. The precaution adoption process model. In: Health Behavior and Health Education: Theory, Research, and Practice, 4th ed. San Francisco: Jossey-Bass; 2008:123-147.

13.   Carter-Harris L, Brandzel S, Wernli KJ, Roth JA, Buist DSM. A qualitative study exploring why individuals opt out of lung cancer screening. Fam Pract 2017 Dec 01;34(2):239-244. [doi: 10.1093/fampra/cmw146] [Medline: 28122849]

14.   Dharod A, Bellinger C, Foley K, Case LD, Miller D. The reach and feasibility of an interactive lung cancer screening decision aid delivered by patient portal. Appl Clin Inform 2019 Jan;10(1):19-27. [doi: 10.1055/s-0038-1676807] [Medline: 30625501]

15.   Enders C. Applied Missing Data Analysis. New York, NY: Guilford Press; 2010.

16.   Lau YK, Caverly TJ, Cherng ST, Cao P, West M, Arenberg D, et al. Development and validation of a personalized, web-based decision aid for lung cancer screening using mixed methods: a study protocol. JMIR Res Protoc 2014 Dec 19;3(4):e78 [FREE Full text] [doi: 10.2196/resprot.4039] [Medline: 25532218]

17.   Lung cancer screening decision tool. Memorial Sloan Kettering Cancer Center. URL: http://nomograms.mskcc.org/Lung/Screening.aspx [accessed 2020-02-15]

18.   Screening for lung cancer pamphlet. *. URL: http://www.prevention.va.gov/docs/LungCancerScreeningHandout.pdf [accessed 2020-02-15]

19.   Chen Y, Marcus MW, Niaz A, Duffy SW, Field JK. MyLungRisk: a user-friendly, web-based calculator for risk assessment of lung cancer based on the validated Liverpool Lung Project risk prediction model. International Journal of Health Promotion and Education 2014 Feb 25;52(3):144-152. [doi: 10.1080/14635240.2014.888814]

20.   Volk RJ, Linder SK, Leal VB, Rabius V, Cinciripini PM, Kamath GR, et al. Feasibility of a patient decision aid about lung cancer screening with low-dose computed tomography. Prev Med 2014 May;62:60-63 [FREE Full text] [doi: 10.1016/j.ypmed.2014.02.006] [Medline: 24518006]

Abbreviations

SRM: standardized response mean
USPSTF: US Preventive Services Task Force

XSL•FO

RenderX

<u>Original Paper</u>

Balancing Accuracy and Privacy in Federated Queries of Clinical Data Repositories: Algorithm Development and Validation

Yun William Yu[1], PhD; Griffin M Weber[2], MD, PhD

[1]Computer & Mathematical Sciences, University of Toronto, Toronto, ON, Canada
[2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Griffin M Weber, MD, PhD
Department of Biomedical Informatics
Harvard Medical School
10 Shattuck St
Boston, MA, 02115
United States
Phone: 1 617 432 6134
Email: weber@hms.harvard.edu

Abstract

Background: Over the past decade, the emergence of several large federated clinical data networks has enabled researchers to access data on millions of patients at dozens of health care organizations. Typically, queries are broadcast to each of the sites in the network, which then return aggregate counts of the number of matching patients. However, because patients can receive care from multiple sites in the network, simply adding the numbers frequently double counts patients. Various methods such as the use of trusted third parties or secure multiparty computation have been proposed to *link* patient records across sites. However, they either have large trade-offs in accuracy and privacy or are not scalable to large networks.

Objective: This study aims to enable accurate estimates of the number of patients matching a federated query while providing strong guarantees on the amount of protected medical information revealed.

Methods: We introduce a novel probabilistic approach to running federated network queries. It combines an algorithm called HyperLogLog with obfuscation in the form of hashing, masking, and homomorphic encryption. It is *tunable*, in that it allows networks to balance accuracy versus privacy, and it is computationally efficient even for large networks. We built a user-friendly free open-source benchmarking platform to simulate federated queries in large hospital networks. Using this platform, we compare the accuracy, k-anonymity privacy risk (with k=10), and computational runtime of our algorithm with several existing techniques.

Results: In simulated queries matching 1 to 100 million patients in a 100-hospital network, our method was significantly more accurate than adding aggregate counts while maintaining k-anonymity. On average, it required a total of 12 kilobytes of data to be sent to the network hub and added only 5 milliseconds to the overall federated query runtime. This was orders of magnitude better than other approaches, which guaranteed the exact answer.

Conclusions: Using our method, it is possible to run highly accurate federated queries of clinical data repositories that both protect patient privacy and scale to large networks.

Introduction

Background

Widespread adoption of electronic health records has generated vast amounts of data, which are increasingly being used in clinical, epidemiological, and public health research [1]. Data from multiple health care organizations are often needed to increase statistical power or to access diverse patient populations and geographic regions. Although it is possible to combine patient-level data from multiple sites into a secure central repository for analysis, there are often significant technical and regulatory barriers to doing this in a way that ensures patient privacy. Institutions must compare the benefit of centralized data for research with the risk of violating the Health Insurance

Portability and Accountability Act (HIPAA) and other privacy laws as a result of unintended disclosure of patient data. An alternative approach is to create federated clinical data research networks, which broadcast queries to multiple sites, run analyses locally, and then combine the results. In this way, sites retain control over their patient data. Two of the largest networks in the United States are the Patient-Centered Outcomes Research Network (PCORnet) [2] and the National Institutes of Health (NIH)–funded Accrual to Clinical Trials (ACT) network [3-5], both of which connect dozens of health care organizations across the country and include health data on nearly 100 million Americans.

As patients often receive care at more than one clinical site, the data for a patient at any one site might not be complete, and the same information about a patient might be duplicated at different sites. This can lead to queries returning incorrect results. This problem is amplified when the sites in the network are geographically close and there is greater overlap in their patient populations. However, because patients move or travel, sometimes across state or country borders, even far apart sites might share patients. A similar situation arises when patients' data are intentionally separated for technical reasons, such as when large amounts of clinical data (eg, diagnoses and medications) and genomic data are stored in different locations, and it is not feasible to merge them into a single database. In both cases, computation must be performed in a distributed fashion, but the challenge is that an individual patient's data may be spread across multiple databases.

Various methods to addressing this problem have been described in the literature, but they have different trade-offs in terms of accuracy, privacy, scalability, and computational complexity. We grouped these into 3 broad categories: aggregate counts, hashed patient identifiers, and privacy-guaranteed methods (Figure 1).

Figure 1. Federated query methods. We classify methods for merging distributed queries into 3 groups: (top) sharing aggregate counts, (middle) sending full hashed patient identifiers, and (bottom) generating bitstrings (displayed as hexadecimal) that do not directly correspond to individual patients but can be merged together. HLL: HyperLogLog; MPC: multiparty computation; SSN: social security number.
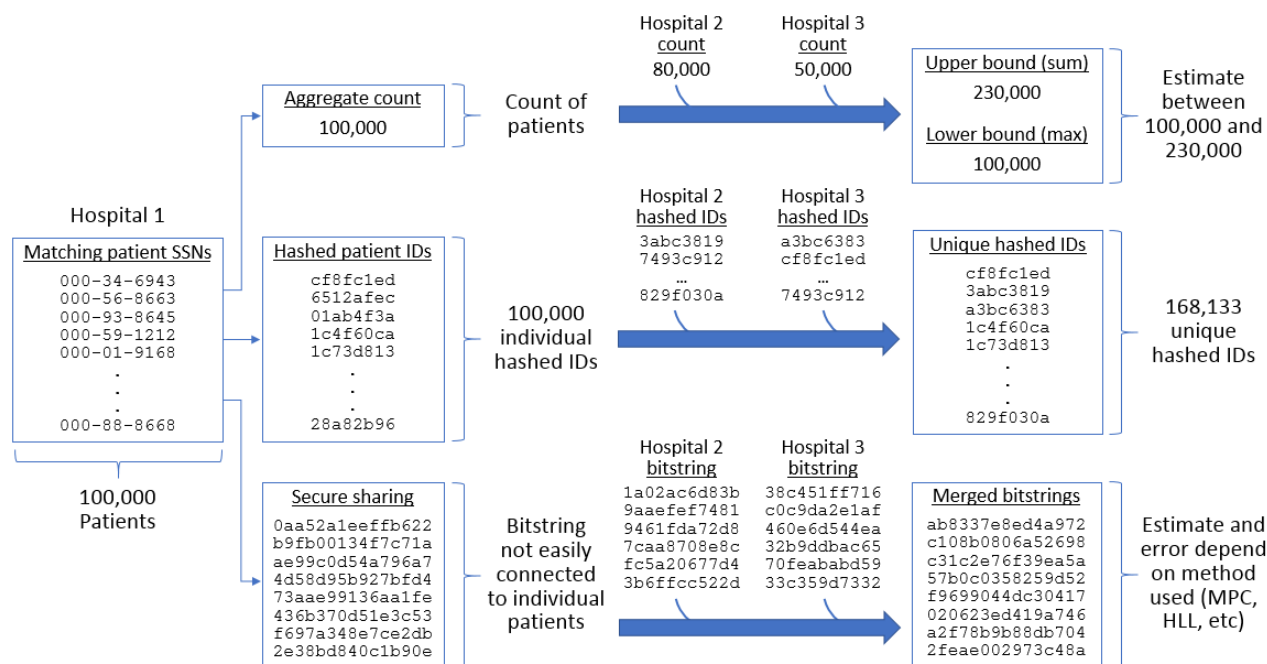
Aggregate Counts

Federated queries in PCORNet and ACT ask sites to return the number of patients in their local databases who match some set of criteria, such as having both hypertension and diabetes. The networks present the user with the aggregate count from each site, and no attempt is made to link patients across sites or deduplicate records. This can lead to large overestimates of the number of distinct patients who match a query if the counts from each site are naively summed [6]. To protect patient privacy, the networks mask small counts by displaying ≤*10 patients*. However, it is possible to combine results from multiple queries to reveal information about individual patients (see the *Methods* section for details). Sites participating in these networks are aware of this privacy risk, which they mitigate through institutional agreements that require sites to audit researchers' queries and monitor their use of the network.

Hashed Patient Identifiers

The most accurate and semisecure method to deduplicate the results in a federated query is for each site to return the full list of patients who match the query. Privacy is the main concern, as data on every patient matching the query (potentially many millions of people) must be shared. Patient identifiers (eg, name and date of birth) [7] are typically encrypted using a one-way hash function, such as Secure Hash Algorithm 1 (SHA-1) [8]. The same patient at two sites will be hashed to the same value if the same hash function is used (and there are no inconsistencies in the underlying demographic data). Unfortunately, hash functions are vulnerable to dictionary or linkage attacks, where an adversary who knows the encryption

method can simply generate a rainbow table of the hashes of many possible patient identifiers (eg, exhaustively searching all 9-digit social security numbers or accessing public voter registration lists) and then use this to reidentify the list of hash values returned by a site [9].

Privacy-Guaranteed Methods

Secure multiparty computation (MPC) and homomorphic encryption techniques enable true privacy guarantees in a federated network (see the *Methods* section) and have recently been introduced for distributed genome-wide association studies [10] and pharmacological collaboration [11]. The limitation of these algorithms is their computational complexity. Protocols that securely determine the number of shared patients between two sites [6,12-14] are impractical for large networks because the number of pairwise and multiway comparisons grows exponentially with the number of sites. Other approaches that avoid exponential comparison either require sharing gigabytes of data [15], making numerous rounds of back-and-forth communication [16], or using trusted third parties [17]. These are also problematic because, as we have previously shown [18], large federated clinical data networks are fragile, with multiple sites typically failing to respond even to aggregate count queries.

HyperLogLog Sketch

In this paper, we propose a new method for combining data from sites in a federated clinical data network, based on the HyperLogLog (HLL) probabilistic sketching algorithm [19]. A probabilistic sketch is a small data structure that summarizes large amounts of data. A calculation can run on the sketch to obtain a fast, accurate estimate of what the result would be on the original data. Although HLL is widely used in many software programs, such as internet search engines, to our knowledge, it has not been applied to federated queries of health data.

The basic idea behind HLL (and other minimum value sketches) [20] is that the minimum of a collection of random numbers between 0 and 1 is inversely proportional to how many numbers are present. For example, a single random number between 0 and 1 has an expected value of 0.5; however, if we have 99 random numbers, the minimum has an expected value of 0.01. By using a hash function that maps patients to a random number between 0 and 1, we can estimate the number of patients who match a query at a site by keeping track of just the minimum hash value of the matching patients. If the minimum hash value is v, then the estimated number of patients is $(1/v)$-1. Although the accuracy of this estimate is poor, the method can be improved by using t different hash functions to generate t independent estimates of the number of patients. The average of these results in a more accurate overall estimate. The set of t minimum hash values is the sketch.

If each site in a network uses the same hash function and returns its minimal hash value, then we can estimate the number of distinct patients in the whole network that match the query from the smallest of those values. Although it may seem unintuitive that the network minimum hash is the same as the hash for one hospital, the hospital which the minimum hash corresponds to

changes when multiple hash functions are used, allowing the estimator to be accurate.

Instead of using t hash functions, HLL improves the accuracy of this method by using a single hash function but efficiently dividing the patients into t partitions and returning the minimum hash value of patients in each partition. HLL also returns the position of the leading one indicator in the binary expansion of the minimum values rather than the actual values. This only has a small effect on accuracy; however, it greatly reduces the risk of reidentification from a dictionary attack. For t partitions, the relative error of HLL is approximately $1/sqrt(t)$. For example, by asking sites to share an HLL sketch with only 100 values, the number of distinct patients can be estimated with a 10% relative error. The error can be reduced by increasing t. Although *higher t* increases the risk of reidentification, the risk is quantifiable and predictable, enabling networks to define policies that maximize accuracy while reducing risk to an acceptable level.

Objectives

We aim to enable accurate estimates of the number of unique patients matching a federated query while providing strong guarantees on the amount of protected medical information revealed.

Structure of This Paper

In the *Methods* section, we first show how sites can generate a privacy-preserving HLL sketch of the patients who match a query and how the shared sketches from sites can be combined to estimate the number of unique patients in the network who match the query. We then describe several *obfuscation* approaches that further reduce the privacy risk of aggregate counts, hashed identifiers, and HLL sketches. These include methods that might result in a loss of information or an increase in computational complexity to make it more difficult or impossible for an adversary to identify patients. In the *Results* section, we test our algorithm and other methods using simulated networks of different sizes and degrees of patient overlap. We compare them along several dimensions, including accuracy, privacy risk, computation time, and amount of data shared. Finally, in the *Discussion* section, we summarize the trade-offs and limitations of the algorithms and provide recommendations on when networks should consider using HLL sketches.

Methods

Algorithms and Obfuscation Techniques for Federated Queries

Here, we describe the algorithms we compared. The basic model assumes that a researcher at one hospital in the network sends a query of the form *How many unique patients have condition X across the hospital network?* to a central network hub. The hub then distributes the query to all the hospitals in the network. The hospitals determine which of their patients match the query and return a result (the form of this result varies by algorithm) to the hub. The hub combines the results and returns an estimate of the total number of unique patients to the researcher. The name of each algorithm combines the base method (*Count*,

HashedIDs, or *HLL*) and any additional obfuscation (*Mask*, *MPC*, *Rehash*, or *Shuffle*).

Count

Each hospital runs the researcher's query locally and sends the hub a single count of the number of matching patients. The hub returns 2 numbers: (1) the maximum count from a hospital and (2) the sum of counts from all hospitals. The maximum count corresponds to a lower bound on the result, because even in the event of significant overlapping patients between hospitals, there are at least as many unique patients across the network as there are at a single hospital. For example, in Figure 1, hospitals 1, 2, and 3 have 100,000, 80,000, and 50,000 patients, respectively. It might be the case that all patients at hospitals 2 and 3 are also patients at hospital 1, which has the maximum count. However, this is not possible for the hospitals with smaller counts. For example, out of 100,000 patients of hospital 1, at most 80,000 can also be patients at hospital 2. The sum of all counts is obviously an upper bound, although it might be a substantial overestimate when there is a significant overlap between hospitals. Conversely, the maximum of all counts is obviously a lower bound.

Count+Mask

The procedure is identical to Count, except that if the actual count of a hospital is between 1 through 9 inclusive, the hospital returns 10 to the hub instead. This masking procedure ensures that no nonzero number corresponds to fewer than 10 patients, ensuring 10-anonymity. Both the PCORNet and ACT networks use Count+Mask. ACT further obfuscates the result by adding a small random number between –10 and +10 to the actual count [4]; however, we ignore this in our analyses.

Count+MPC

This protocol is based on the ElGamal cryptosystem [21] using a distributed private key to ensure that no one party can decrypt intermediate data. Only the final sum is decrypted. The individual hospital counts are hidden, even if all hospitals but one and the hub are compromised. The major disadvantage is that the MPC requires all hospitals to respond before any answer can be given. In large networks, it is likely that some hospitals will either be slow to respond or not respond at all [18], which limits this protocol to only small networks in practice (for additional information on our MPC implementation, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]).

HashedIDs

Each hospital runs the query locally, producing a list of matching patient IDs. Each hospital needs to use the same process for constructing IDs so that the same patient at different hospitals will have the same ID. As there is no universal patient identifier, the ID should be based on information likely to be unique to the patient and available at all hospitals, such as the concatenation of the patient's first name, last name, and date of birth [7] (for additional details and limitations of generating a patient ID, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]). Patient IDs are encrypted using a one-way hash function. For our simulations, we used SHA-1, but in practice, a newer, more secure hash function should be used. The list of hashed IDs is then sent back to the hub. The hub then counts the number of distinct hashed IDs received from all sites and returns this as the exact answer to the query. Sites can precompute the hashed IDs for all of their patients to improve the performance of queries. Note that because *HashedIDs* uses the same hash function for all queries, a dictionary or linkage attack by the hub has a high likelihood of success.

HashedIDs+Rehash

This is identical to *HashedIDs*, except that the originating hospital (the hospital with the researcher who ran the query) also sends the hub a random string encrypted with the public keys of each of the other hospitals (using any kind of standard off-the-shelf asymmetric key encryption, as used in protocols such as Rivest-Shamir-Adleman [RSA] and Hypertext Transfer Protocol Secure [HTTPS]). Each hospital rehashes all the patients, prepending the random string before running it through SHA-1. By doing so, because the hub does not know the random prefix string, it cannot perform a dictionary attack to reverse the hash function, and thus, all patients get 10-anonymity. Of course, rehashing all patients with each query requires additional computational time.

HLL

A graphical overview of HLL is shown in Figure 2. Like HashedIDs, in HLL, the hospital uses the SHA-1 hash function to produce a 160-bit pseudorandom number for each patient that matches a query. The first 64 bits are interpreted as an integer B, and the patient is put into bucket B mod t, where t is the number of buckets. The hospital then finds the position V of the first bit set to 1 in bits 65 to 128 of the SHA-1 string. Within each bucket, the hospital stores the largest value V corresponding to a patient. The list of bucket values is the HLL sketch from that hospital. (Note that like *HashedIDs*, hospitals can precompute the buckets B and values V for all of their patients, so that this step does not have to be repeated for each query.)

Figure 2. HLL sketches. (a) To create an HLL sketch, we first hash a set of identifiers for the matching patients (eg, social security number) to binary strings. The first several bits of each binary string are used to bucket the values, and then within each bucket, we store the position of the leading one indicator of the minimum value. (b) HLL sketches from different hospitals are merged by simply taking, within each bucket, the maximum value across sketches. (c) Given a list of buckets, we can estimate the cardinality. HLL: HyperLogLog; SSN: social security number.

(a) Creating an HLL sketch
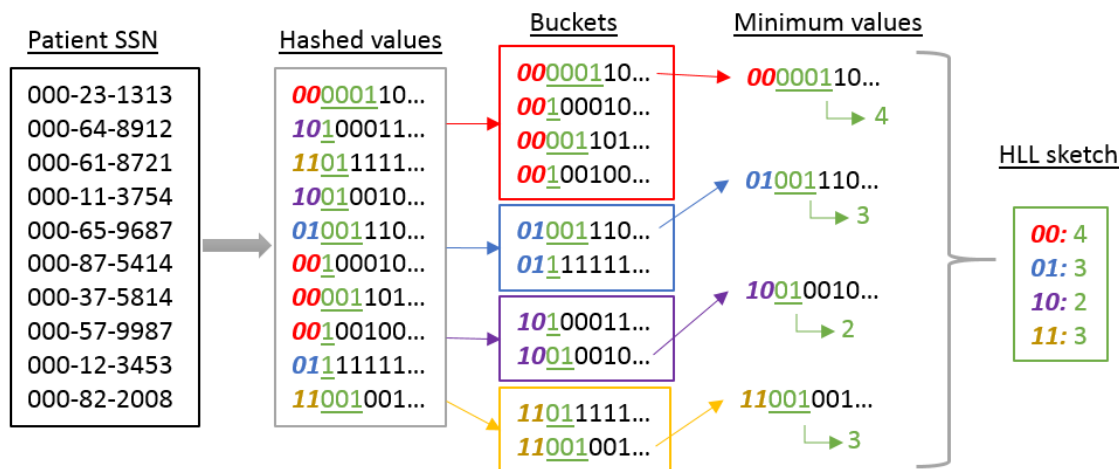
(b) Merging HLL sketches
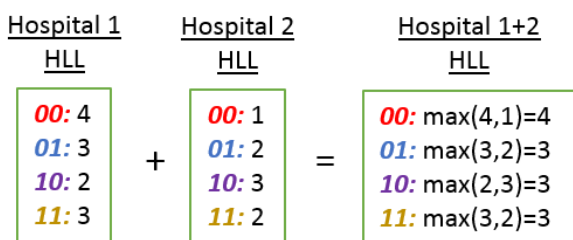
(c) Estimating cardinality

$$E = \left(t \int_0^\infty \left(\log_2 \left(\frac{2+u}{1+u} \right) \right)^t du \right)^{-1} \cdot t^2 \cdot \left(\sum_{i=1}^m 2^{-T[i]} \right)^{-1}$$
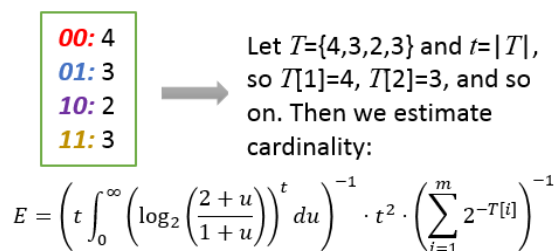
The hospitals send these HLL sketches to the central hub. The hub combines the sketches by taking the maximum within each bucket across the hospital sketches, generating a sketch of the union. The hub then estimates the cardinality C of the union sketch using the standard HLL estimator [19]. The hub also provides a 95% CI by using the fact that the SD of the estimate is around 1/sqrt(t), so 1±1.96/sqrt(t) gives the lower and upper bounds of a 95% CI.

HLL+Mask

As shown in Figure 3, this algorithm is identical to HLL, except that the hospital precomputes a list of bucket values that are less than 10-anonymous. If after generating the HLL sketch corresponding to the query, a hospital sees that there is a bucket that is not 10-anonymous, the hospital aborts and reverts to the *Count+Mask* algorithm, where only a single (possibly masked) aggregate count is returned. The hub thus receives a combination of sketches and masked counts.

Figure 3. Applying obfuscation to HLL sketches. (a) HLL+Mask: For each bucket, we count the total number of patients (not just the ones who match the query) whose hashes have the same leading 1-indicator. If that number is less than 10, then the bucket is not 10-anonymous, so we do not send the HLL sketch. Instead, we only send a masked aggregate count of the number of patients matching the query. (b) HLL+Shuffle: We do a coordinated random shuffling so the central hub does not know what the original buckets were for the leading 1 indicator. However, the hub can still estimate cardinality in the same way as HLL without obfuscation. HLL: HyperLogLog.
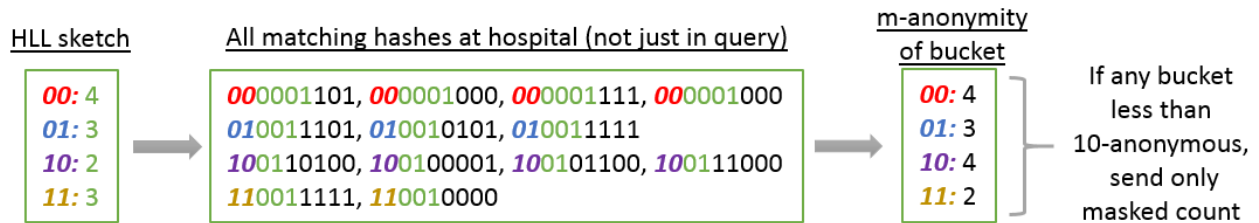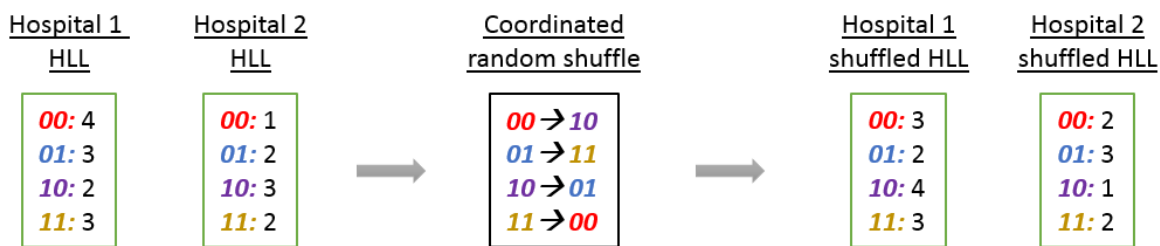
The hub combines the sketches together using the HLL cardinality estimator to obtain an estimate of the count of the union of all the hospitals that sent sketches with appropriate 95% error bounds. From that, the hub goes through something similar to *Count*. The hub returns 2 numbers: the sum of all raw hospital counts plus the 95% CI maximum for the HLL union count, which gives an upper bound, and the maximum of the set of raw counts or the 95% CI minimum for the HLL union, which gives a lower bound.

HLL+Rehash

This algorithm uses HLL but with an obfuscation method similar to *HashedID+Rehash*. When the originating hospital sends a query to the hub, it sends both a query and a random string encrypted with public keys of each of the other hospitals in the network. The hospitals completely regenerate the HLL sketch while prepending the random string to the patient IDs before hashing. Although this procedure takes more time, the hub cannot use a dictionary attack at all because it does not know the random string. Thus, all patients are guaranteed 10-anonymity if the random string is not revealed to the hub.

HLL+Shuffle

This algorithm also sends a random string encrypted with public keys of each of the other hospitals in the network to the hub. However, it is much faster than *HLL+Rehash* because it avoids having to rehash all patients. Each hospital first creates an ordinary HLL sketch using their precomputed hashed IDs. It then shuffles the ordering of the buckets using the random string to determine the sort order and then sends the shuffled sketch to the hub ([Figure 3](#)).

As every hospital uses the same permutation, the sketches can still be combined and the normal estimators can be used.

However, the hub, without knowing the random string, cannot know which bucket in the original sketch corresponds to a bucket in the shuffled sketch. Normally, an HLL bucket is less than 10-anonymous if that value+bucket pair corresponds to fewer than 10 individuals at the hospital. With shuffling, an HLL bucket is less than 10-anonymous only if that value corresponds to fewer than 10 individuals at the hospital. On average, this decreases the risk by dividing the risk score by the number of buckets. In other words, the buckets partition the patient population into smaller, more identifiable groups. By shuffling the buckets, it is no longer known which partition the value came from, which makes the value less identifiable.

HLL+MPC

Like *Count+MPC*, this method is based on the ElGamal homomorphic cryptosystem, and we use the same primitives as in that method (with the same security guarantees). We additionally take inspiration from a previous paper applying MPC to a Flajolet-Martin style approximate counter [16]. The key setup, exchange, encryption and decryption routines are identical to those of *Count+MPC* (for additional information on our MPC implementation, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]).

HLL+Shuffle+MPC

This procedure is simply a combination of *HLL+Shuffle* and *HLL+MPC*. Each hospital simply shuffles their buckets according to the random string before performing the encryption. The rest of the procedure is identical to that of *HLL+MPC*.

Testing and Evaluating the Algorithms

To quantitatively measure privacy loss, we used an adapted *k*-anonymity model of privacy, whereby the privacy risk is defined to be the number of revealed data points that correspond

to fewer than $k=10$ patients [22,25] (for details on the privacy risk score, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]). We ran benchmarks for runtime, accuracy, and privacy loss on (1) shared aggregate counts (*Count* and *Count+Mask*), (2) shared hashed identifiers (*HashedIDs*), and (3) our proposed HLL approach. Each of these was paired with various obfuscation techniques of masking, rehashing, shuffling, and MPC. HLL was tested using different number of buckets or values in the sketch. We indicate the size of the sketch, *t*, with a number after *HLL*, such that *HLLN* means 2^N values. For example, $t=2^1=2$ (*HLL1*), $t=2^4=16$ (*HLL4*), $t=2^7=128$ (*HLL7*), and $t=2^{15}=32,768$ (*HLL15*). Although *Count+MPC* uses a standard MPC privacy-guaranteed cryptosystem, we implemented our own protocols for the HLL+MPC variants using ElGamal encryption [21] and a private equality test [23]. We did not run benchmarks for other existing privacy-guaranteed methods because they do not scale well and are infeasible for running on large data sets, with either extremely high runtime or error (for descriptions of several of these algorithms and their limitations, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]).

Due to patient privacy, we cannot test the algorithms using actual hospital data. Therefore, we developed software for generating simulated federated networks of hospitals spread geographically with highly varying sizes and overlap [24] (for details on simulating a federated hospital network, see Multimedia Appendix 1 [6,7,10-12,14-17,21-24]). We ran our benchmarks on simulated networks containing up to 100 million total distinct patients, distributed across 100 hospitals. In the simulations, patients on average received care at 2 hospitals. However, this number varies and hospitals that are geographically close in the simulations are modeled to have a larger number of shared patients.

The benchmarks were run on an 8-core AMD Ryzen 1700 processor with 16 GB of RAM running Ubuntu 18.04.2 Long Term Support. We measured the wall-clock time for each pipeline component for time complexity and serialized bitstrings in each communication round for transmission space complexity. We provide all code in GitHub [26].

Results

Quantitative Simulation Benchmark Results

Multimedia Appendix 2 lists the detailed benchmark results for accuracy, privacy risk, and runtimes of queries matching 1, 10, 100, 1000, 10,000, 100,000, 1 million, 10 million, or 100 million patients using the different methods. As an example, Table 1 shows a subset of rows from the table in Multimedia Appendix 2 corresponding only to queries matching 10,000 patients and HLL sketches with 2^7 (HLL7) and 2^{15} (HLL15) values.

Accuracy is described in absolute terms as the 95% CIs of the estimated number of patients who matched a query in 100 simulated experiments. More precisely, in each of the 100 runs, each estimator tries to return either its best guess or upper or lower bounds. If it returns a single best guess, then we report the 97.5 and 2.5 percentiles as the upper and lower bounds, respectively. If it returns upper or lower bounds, then we report the 97.5 percentile of the upper bound and the 2.5 percentile of the lower bound. These are then converted into relative errors by comparing them with the true number of distinct patients.

Privacy risk is determined by counting the number of statistics (ie, a count, HLL bucket, or hash) that are not 10-anonymous revealed to either the hub or the hub colluding with a hospital. It relates to the number of patients who are *potentially* identifiable with a specific statistic, but it does not necessarily mean that an adversary will be able to identify a patient from a statistic. Therefore, it can be thought of as an upper bound on direct linkage risk. Note that this guarantee is applicable primarily for one common threat model. In the *Discussion* section, we will cover some other more sophisticated potential avenues for attack.

Wait time is the additional computational time that hospitals require to generate the statistics plus the time the hub requires to combine each hospital's results. (It does not include the time each hospital needs to run the query.) For the same query, hospitals might have different wait times based on the number of matching patients. We, therefore, report both *mean wait time*, which is the average hospital computation time+hub computation time, and *max wait time*, which is the maximum hospital computation time for a run+hub computation time.

Table 1. Benchmark results for selected methods for queries matching 10,000 patients.

Method and obfuscation	Estimated number of patients		Wait (seconds)		Risk:Hub	Risk: Hub+Site
	Range of counts	Relative error (%)	Mean	Max		
Count						
None	899.9-19,470	–91 to 95	0	0	2.65	2.65
Mask	899.9-19,477	–91 to 95	0	0	0	0
MPC[a]	18,886-19,470	89 to 95	0.099	0.099	0	0
HLL7[b]						
None	8310-11,347	–17 to 13	0.006	0.006	15.73	15.73
Shuffle	8310-11,347	–17 to 13	0.006	0.006	0.23	15.73
Rehash	8310-11,347	–17 to 13	0.007	0.016	0	15.73
Mask	7167-14,123	–28 to 41	0.005	0.005	0	0
MPC	8310-11,347	–17 to 13	37.83	37.83	0.3	0.3
Shuffle+MPC	8310-11,347	–17 to 13	37.83	37.83	0	0.3
HLL15						
None	9928-10,075	–1 to 1	1.462	1.462	3707	3707
Shuffle	9928-10,075	–1 to 1	1.462	1.462	0.23	3707
Rehash	9928-10,075	–1 to 1	1.625	1.668	0	3707
Mask	899.9-19,477	–91 to 95	0.012	0.012	0	0
HashedIDs						
None	10,000-10,000	0 to 0	0.002	0.002	19,174	19,174
Rehash	10,000-10,000	0 to 0	0.002	0.004	0	19,174

[a]MPC: multiparty computation.

[b]HLL: HyperLogLog.

As an example, in Table 1, for a query that actually matches 10,000 patients, the basic *Count* algorithm had an estimated count CI (using the summation for the upper estimate and maximum for the lower estimate) of 899.9 to 19,470 patients or a relative error of –91% to +95%. It also, on average, had 2.65 hospitals that returned potentially identifiable counts because the value was less than 10. This risk can be eliminated with *Count+Mask*, which increases the error, or by *Count+MPC*, which adds computational complexity, and only gives a single guess, instead of both upper and lower bounds. On the opposite extreme, *HashedIDs* returns the exact answer, but all 10,000 patients' identities are at risk from a dictionary attack. (Note that Table 1 lists the risk for *HashedIDs* at 19,174 because the same patient's hash value can be returned by more than one hospital. We report the number of potentially identifiable values shared, not the number of unique patients at risk.) In *HashedIDs+Rehash*, the hub alone cannot identify patients from the hash values (the *Risk:Hub* column). However, the risk returns if an adversary can also obtain the secret random string from a hospital (the *Risk:Hub+Site* column).

Table 1 shows that *HLL7* and *HLL15* can achieve a more tunable balance between accuracy and privacy. *HLL7* has a relative error of –17% to +13% (8310 to 11,347), which is considerably better than that of *Count*, and *HLL15* results in an even smaller relative error of –1% to 1% (9928 to 10,075). *HLL7* and *HLL15* generate, on average, 15.73 and 3707 potentially identifiable values. However, adding obfuscation with *HLL+Shuffle* adds essentially no additional computation time but reduces the risk to less than 1 (0.23 on average) potentially identifiable value. In other words, highly accurate estimates with only 1% error can be obtained with most queries having no risk of reidentification. Even if an adversary obtains the secret random string, the risk of 3707 is much less than 19,174 for *HashedIDs*.

Graphical Comparison of Algorithms

Figure 4 graphically illustrates the accuracy (the horizontal axis) and risk (the vertical axis) trade-off of the different algorithms. For simplicity, only the upper bound of the relative error is used for accuracy. (The lower bound and absolute errors are not shown.) Although an individual simulation is plotted as a single point in the figure, algorithms are shown as regions because changing the input parameters to the simulation affects the results. For example, the blue region in Figure 4 covers the range of HLLs with queries of different sizes (10 to 10 million matching patients) and sketches of different sizes (HLL1=2 to HLL15=32,768 values).

Figure 4. Comparison of the query accuracy/privacy risk trade-off based on the simulations of a network with 100 sites and 100 million patients. HashedIDs and Count bound the graph, whereas HLL-based methods enable a more balanced approach. (HLL+MPC is only shown for 10 million patients, and the values for HLL7+MPC and HLL15+MPC are theoretical rather than experimental.) HLL+MPC reduces the HLL risk by 1/s, where s is the number of sites in the network. HLL+Shuffle reduces the HLL risk by 1/t, where t is the number of values in the HLL sketch. HLL: HyperLogLog; MPC: multiparty computation.
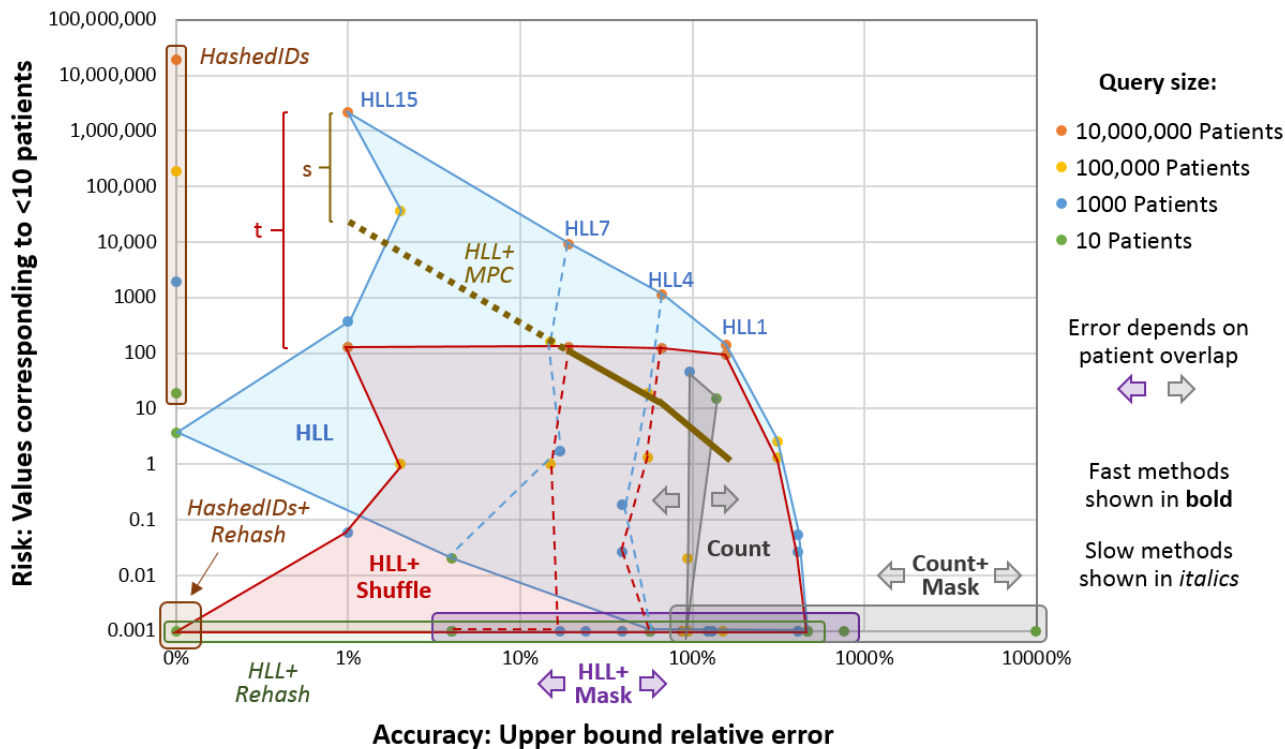
The key takeaway from Figure 4 is that *Count* and *HashedIDs* are extremes that cover only one axis or the other, whereas variations of *HLL* enable networks to select an algorithm that fits anywhere between the axes. In other words, with *HLL*, networks can determine an acceptable risk level and pick the sketch size and obfuscation method that will give the most accurate result. Alternatively, they can start with a desired accuracy and pick the most secure method that runs within a given amount of time.

Count+Mask has the worst accuracy but guarantees 10-anonymity (thin horizontal gray box; Figure 4). As each patient in the simulation was, on average, at two hospitals, queries that matched all 100 million distinct patients returned counts from each hospital that added up to 200 million—a 100% overestimate. Queries that only matched a few patients (*small queries*) had much greater error because of the obfuscation. The worst case, in theory, is when a query matches one distinct patient and that patient happens to be at each of the 100 hospitals. As each hospital returns ≤10, the upper bound estimate assumes that there are 10 patients in each hospital and that there is no overlap. This would result in an upper bound estimate of 100×10=1000 or a relative error of 99,900%. Even when patients are only at one hospital (no overlap), *Count+Mask* can have a 900% error.

Without obfuscation, the relative error of *Count* in the simulations remained near 100% for queries of all sizes (thin vertical gray box; Figure 4). However, for small queries, many sites returned potentially identifiable counts less than 10. At the other extreme, *HashedIDs* always gave correct answers (0%

relative error). However, this requires sharing individual data on all matching patients (thin vertical brown box; Figure 4). The risk can be reduced if a different hash function is used for each query (*HashedIDs+Rehash*) and an adversary is unable to discover the hash functions.

Variations of *HLL* fill in the space between *Count*, *Count+Mask*, and *HashedIDs*, allowing the networks to tune their estimation method to achieve a more desirable balance of accuracy and risk for a given application. In Figure 4, *HLL* (the blue region), *HLL+Shuffle* (the red region), and *HLL+Rehash* (the thin horizontal green box) have the same accuracy but different levels of risk. In contrast to *Count*, which has more risk with smaller queries, *HLL*, like *HashedIDs*, has a higher risk with larger queries. Doubling the number of buckets in the *HLL* sketch reduces the error by a factor of sqrt(2); however, without obfuscation, it also doubles the risk.

The benefit of *HLL+Shuffle* is that buckets can be added to reduce error with only minimal change in risk. For queries that matched fewer than 100,000 patients, even *HLL15+Shuffle*, which has a relative error of only approximately 1%, had an average privacy risk of less than 1. *HLL+Rehash* reduced risk even further but required over a minute of extra computational time in some experiments, whereas the computational time of HLL+Shuffle is negligible. *HLL+Mask* guarantees 10-anonymity, but its error was often almost as large as *Count+Mask*. The benefit of *HLL+Mask* is that it can leverage the improved accuracy of HLL when possible, while ensuring that no added risk is introduced.

Qualitative Comparison of the Algorithms

Table 2 provides a qualitative summary of the results. In general, HLL, especially with obfuscation, is much more accurate than aggregate counts, lower risk than sharing hash values of all matching patients, and more scalable than privacy guaranteeing

algorithms. The relevant benefits of certain methods depend on the number of patients who match the query. For example, as the number of patients increases, the risk of *Count* decreases, as indicated by "(–)", while the risk of *HLL7* increases, as indicated by "(+)."

Table 2. Qualitative comparison of algorithms.

Method and obfuscation	Approximation error	Runtime wait	Risk:Hub	Risk:Hub+Site
Count				
None	Large	Very small	Medium (–)	Medium (–)
Mask	Large	Very small	Zero	Zero
MPC[a]	No change[b]	Medium	Zero	Zero
HLL7[c]				
None	Medium	Small	Medium (+)	Medium (+)
Shuffle	No change	No change	Small (+)	No change
Rehash	No change	Medium (+)	Zero	No change
Mask	Medium (+)	Medium (–)	Zero	Zero
MPC	No change	Large	Small (+)	Small (+)
Shuffle+MPC	No change	HLL7+MPC	Very small (+)	HLL7+MPC
HLL15				
None	Small	Medium	Large (+)	Large (+)
Shuffle	No change	No change	Small (+)	No change
Rehash	No change	Medium (+)	Zero	No change
Mask	Large (+)	Medium (–)	Zero	Zero
HashedIDs				
None	Zero	Medium (+)	Very large (+)	Very large (+)
Rehash	No change	No change	Zero	No change

[a]MPC: multiparty computation.

[b]No change: the value is the same as the method without any obfuscation.

[c]HLL: HyperLogLog.

Computational and Communication Costs

Multimedia Appendix 3 shows the theoretical upper bounds on the computational costs of each method plus obfuscation technique, theoretical exact communication costs (the space complexity of the amount of data that the hospitals and hub have to send over the network), and the actual empirical results of both computational and communication costs.

Discussion

Summary of Results and Practical Considerations

In this study, we surveyed and benchmarked a range of methods for determining the number of distinct patients who matched a federated query, exploring the trade-offs in accuracy, privacy, and speed. We explicitly do not endorse a single one-size-fits-all method because different networks and institutions will have different needs. With data use agreements and a trusted third party, *HashedIDs* provides the most accurate results. When minimizing privacy risk is the most important factor, networks

can choose between (1) fast but inaccurate methods such as *Count+Mask*, (2) accurate but slow algorithms such as *HLL+Rehash*, or (3) privacy-guaranteed methods that only work on small networks. A key goal of the ACT network is *real-time* queries that enable rapid exploration of the data. As a result, adding even a few seconds of computational time to ACT queries might not be acceptable. When runtimes must be minimized, methods such as *HLL7+Mask* and *HLL7+Shuffle* are fast and have a good balance between accuracy and privacy.

In practice, we envision a combination approach. Queries can first be run using a fast, private method, such as *Count+Mask* or *Count+MPC*. Given these rough results and the needs of the researcher, hospitals can then be asked to return the HLL sketches for the patients who matched the query. The initial count estimate and the privacy risk allowed by the network could be used to select the HLL sketch size and obfuscation method that would return the most accurate result in a reasonable amount of time. In the final stage of research (eg, in preparation for a full clinical trial), investigators could request permission

XSL•FO
RenderX

from institutions to run accurate but potentially identifiable queries, such as *HLL15* or *HashedIDs*.

Limitations

It is important for each institution to assess their own risk models. In particular, our risk model assumes that given a sketch for a given condition (eg, hypertension), the adversary already has access to the list of patients at the hospital and wants to identify patients that have the condition. The filled buckets of an HLL sketch correspond to hashes of patients who have the condition, and our goal is to ensure that for every patient with the condition, at least nine other patients without that condition could have hashed to the same value, ensuring 10-anonymity. Statistics that do not meet this requirement count for the privacy loss score. For example, our privacy risk analysis differs considerably from that of Desfontaines et al [27] who argue that "cardinality estimators do not preserve privacy." However, their threat model assumes that an adversary can access the sketches as they are being generated, one patient at a time. In contrast, our risk model is based on each hospital's final sketch, which represents all patients who match the query.

In addition, some amount of information is leaked about the patients *not* included in the sketch, precisely because they were not included. This does not allow an adversary to pinpoint patients with a condition but may sometimes allow them to determine a patient lacking that condition. Of course, this type of leakage is to some extent a problem with any aggregate query system, because if an adversary learns that only 1% of patients at a hospital have a condition, then they know with high certainty that most patients do not. In line with our analysis mentioned earlier, however, for this type of leakage, *Count* is more private than *HLL*, which is more private than *HashedIDs*, so the same privacy-accuracy trade-off applies.

We only considered a federated or distributed network in which no patient-level clinical data leave the institution and queries only return aggregate counts. This is in contrast to privacy-preserving record linkage approaches whose goal is to assemble a centralized deduplicated limited or deidentified data set through an honest broker without exchanging identifiable information. With the appropriate technologies, a secure infrastructure, and the proper institutional agreements in place, it is possible to merge data sets, even on large scales. PCORNet, in particular, has used methods similar to *HashedIDs* and *HashedIDs+Rehash* to do this for subsets of hospitals in its network [28,29]. There are multiple advantages of centralized data, including exact results and ease of use. However, in this study, we showed that (1) linking and deduplicating data at the individual patient level is not necessary to obtain accurate estimates and (2) this can be done in a computationally efficient manner. There are benefits to this federated model. It reduces concerns that hospitals might have in sharing data, it does not require updating and relinking the central database, and it places less dependency on having an honest broker.

Conclusions

We believe that as federated data networks expand to include more institutions and data types (clinical, genomic, environmental, etc), researchers will increasingly depend on fast, accurate, and secure query tools to obtain the greatest possible scientific value from the networks. However, because no single algorithm meets all these requirements, having the ability to select among different methods for a particular application is essential. In this study, we introduce *HLL* and several obfuscation techniques to provide networks with a tunable approach to determine the number of distinct patients who match a query, which is more balanced than commonly used methods that greatly sacrifice accuracy (*Count+Mask*), privacy (*HashedIDs*), or scalability.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details on the algorithms, secure methods that are not scalable to large networks, the privacy risk score, and the federated hospital network simulation.
[PDF File (Adobe PDF File), 149 KB - jmir_v22i11e18735_app1.pdf]

Multimedia Appendix 2
Detailed benchmark results.
[PDF File (Adobe PDF File), 156 KB - jmir_v22i11e18735_app2.pdf]

Multimedia Appendix 3
Time and space complexity for various methods.
[PDF File (Adobe PDF File), 118 KB - jmir_v22i11e18735_app3.pdf]

[XSL•FO](#)

RenderX

References

1. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 2;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]

2. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]

3. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The shared health research information network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc 2009;16(5):624-630 [FREE Full text] [doi: 10.1197/jamia.M3191] [Medline: 19567788]

4. McMurry AJ, Murphy SN, MacFadden D, Weber G, Simons WW, Orechia J, et al. SHRINE: enabling nationally scalable multi-site disease studies. PLoS One 2013;8(3):e55811 [FREE Full text] [doi: 10.1371/journal.pone.0055811] [Medline: 23533569]

5. Visweswaran S, Becich MJ, D'Itri VS, Sendro ER, MacFadden D, Anderson NR, et al. Accrual to clinical trials (ACT): a clinical and translational science award consortium network. JAMIA Open 2018 Oct;1(2):147-152 [FREE Full text] [doi: 10.1093/jamiaopen/ooy033] [Medline: 30474072]

6. Weber GM. Federated queries of clinical data repositories: the sum of the parts does not equal the whole. J Am Med Inform Assoc 2013 Jun;20(e1):e155-e161 [FREE Full text] [doi: 10.1136/amiajnl-2012-001299] [Medline: 23349080]

7. Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. Proc AMIA Symp 2002:305-309 [FREE Full text] [Medline: 12463836]

8. Eastlake D, Jones P. US Secure Hash Algorithm 1 (SHA1). IETF Tools. URL: https://tools.ietf.org/html/rfc3174 [accessed 2020-10-23]

9. Oechslin P. Making a Faster Cryptanalytic Time-memory Trade-Off. In: Annual International Cryptology Conference. 2003 Presented at: CRYPTO'03; August 17-21, 2003; Santa Barbara, CA, USA. [doi: 10.1007/978-3-540-45146-4_36]

10. Cho H, Wu DJ, Berger B. Secure genome-wide association analysis using multiparty computation. Nat Biotechnol 2018 Jul;36(6):547-551 [FREE Full text] [doi: 10.1038/nbt.4108] [Medline: 29734293]

11. Hie B, Cho H, Berger B. Realizing private and practical pharmacological collaboration. Science 2018 Oct 19;362(6412):347-350 [FREE Full text] [doi: 10.1126/science.aat4807] [Medline: 30337410]

12. Kolesnikov V, Matania N, Pinkas B, Rosulek M, Trieu N. Practical Multi-Party Private Set Intersection From Symmetric-Key Techniques. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017 Presented at: CCS'17; October 30-November 3, 2017; Dallas, Texas, USA. [doi: 10.1145/3133956.3134065]

13. de Cristofaro CE, Gasti P, Tsudik G. Fast and Private Computation of Cardinality of Set Intersection and Union. In: International Conference on Cryptology and Network Security. 2012 Presented at: CANS'12; December 12-14, 2012; Darmstadt, Germany. [doi: 10.1007/978-3-642-35404-5_17]

14. Swamidass SJ, Matlock M, Rozenblit L. Securely measuring the overlap between private datasets with cryptosets. PLoS One 2015;10(2):e0117898 [FREE Full text] [doi: 10.1371/journal.pone.0117898] [Medline: 25714898]

15. Fenske E, Mani A, Johnson A, Sherr M. Distributed Measurement With Private Set-Union Cardinality. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017 Presented at: CCS'17; October 30-November 3, 2017; Dallas, Texas, USA. [doi: 10.1145/3133956.3134034]

16. Dong C, Loukides G. Approximating private set union/intersection cardinality with logarithmic complexity. IEEE Trans Inform Forensic Secur 2017 Nov;12(11):2792-2806. [doi: 10.1109/tifs.2017.2721360]

17. Yigzaw KY, Michalas A, Bellika JG. Secure and scalable deduplication of horizontally partitioned health data for privacy-preserving distributed statistical computation. BMC Med Inform Decis Mak 2017 Jan 3;17(1):1 [FREE Full text] [doi: 10.1186/s12911-016-0389-x] [Medline: 28049465]

18. Weber GM. Federated queries of clinical data repositories: scaling to a national network. J Biomed Inform 2015 Jun;55:231-236 [FREE Full text] [doi: 10.1016/j.jbi.2015.04.012] [Medline: 25957825]

19. Flajolet P, Fusy E, Gandouet O, Meunier F. Hyperloglog: the Analysis of a Near-Optimal Cardinality Estimation Algorithm. In: Conference on Analysis of Algorithms. 2007 Presented at: AofA'07; January 1, 2007; Juan des Pins, France p. 1 URL: http://algo.inria.fr/flajolet/Publications/FlFuGaMe07.pdf

20. Bar-Yossef Z, Jayram T, Kumar R, Sivakumar D, Trevisan L. Counting Distinct Elements in a Data Stream. In: International Workshop on Randomization and Approximation Techniques in Computer Science. 2002 Presented at: RANDOM'02; September 13-15, 2002; Cambridge, MA, USA. [doi: 10.1007/3-540-45726-7_1]

21. Elgamal T. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans Inform Theory 1985 Jul;31(4):469-472. [doi: 10.1109/tit.1985.1057074]

22. El Emam K, Dankar F. Protecting privacy using k-anonymity. J Am Med Inform Assoc 2008;15(5):627-637 [FREE Full text] [doi: 10.1197/jamia.M2716] [Medline: 18579830]

23. Jakobsson M, Juels A. Mix and Match: Secure Function Evaluation via Ciphertexts. In: International Conference on the Theory and Application of Cryptology and Information Security. 2000 Presented at: ASIACRYPT'00; December 3-7, 2000; Kyoto, Japan. [doi: 10.1007/3-540-44448-3_13]

XSL•FO
RenderX

24.   Berry BJ. City size distributions and economic development. Econ Dev Cult Change 1961 Jul;9(4, Part 1):573-588. [doi: 10.1086/449923]

25.   Sweeney L. K-anonymity: a model for protecting privacy. Int J Unc Fuzz Knowl Based Syst 2012 May 2;10(5):557-570. [doi: 10.1142/s0218488502001648]

26.   yunwilliamyu / secure-distributed-union-cardinality. GitHub. URL: https://github.com/yunwilliamyu/ secure-distributed-union-cardinality [accessed 2020-10-23]

27.   Desfontaines D, Lochbihler A, Basin D. Cardinality estimators do not preserve privacy. ArXiv 2020:- epub ahead of print. [doi: 10.2478/popets-2019-0018]

28.   Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. J Am Med Inform Assoc 2015 Sep;22(5):1072-1080 [FREE Full text] [doi: 10.1093/jamia/ocv038] [Medline: 26104741]

29.   Bian J, Loiacono A, Sura A, Mendoza VT, Lipori G, Guo Y, et al. Implementing a hash-based privacy-preserving record linkage tool in the oneFlorida clinical research network. JAMIA Open Sep 2019;2(4):562-569. [doi: 10.1093/jamiaopen/ooz050]

Abbreviations

ACT: Accrual to Clinical Trials
HLL: HyperLogLog
MPC: multiparty computation
NIH: National Institutes of Health
PCORnet: Patient-Centered Outcomes Research Network

XSL•FO
RenderX

Original Paper

Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study

Sahil Sandhu[1], BS; Anthony L Lin[2], MD; Nathan Brajer[2], MBA, MD; Jessica Sperling[3], PhD; William Ratliff[4], MBA; Armando D Bedoya[5], MD; Suresh Balu[4], MBA, MS; Cara O'Brien[6], MD; Mark P Sendak[4], MPP, MD

[1]Trinity College of Arts & Sciences, Duke University, Durham, NC, United States

[2]Duke University School of Medicine, Durham, NC, United States

[3]Social Science Research Institute, Duke University, Durham, NC, United States

[4]Duke Institute for Health Innovation, Durham, NC, United States

[5]Division of Pulmonary, Allergy, and Critical Care Medicine, Duke University School of Medicine, Durham, NC, United States

[6]Department of Medicine, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Sahil Sandhu, BS
Trinity College of Arts & Sciences
Duke University
420 Chapel Drive, Box 98033
Durham, NC
United States
Phone: 1 4085504752
Email: sahilssandhu@gmail.com

Abstract

Background: Machine learning models have the potential to improve diagnostic accuracy and management of acute conditions. Despite growing efforts to evaluate and validate such models, little is known about how to best translate and implement these products as part of routine clinical care.

Objective: This study aims to explore the factors influencing the integration of a machine learning sepsis early warning system (*Sepsis Watch*) into clinical workflows.

Methods: We conducted semistructured interviews with 15 frontline emergency department physicians and rapid response team nurses who participated in the Sepsis Watch quality improvement initiative. Interviews were audio recorded and transcribed. We used a modified grounded theory approach to identify key themes and analyze qualitative data.

Results: A total of 3 dominant themes emerged: perceived utility and trust, implementation of Sepsis Watch processes, and workforce considerations. Participants described their unfamiliarity with machine learning models. As a result, clinician trust was influenced by the perceived accuracy and utility of the model from personal program experience. Implementation of Sepsis Watch was facilitated by the easy-to-use tablet application and communication strategies that were developed by nurses to share model outputs with physicians. Barriers included the flow of information among clinicians and gaps in knowledge about the model itself and broader workflow processes.

Conclusions: This study generated insights into how frontline clinicians perceived machine learning models and the barriers to integrating them into clinical workflows. These findings can inform future efforts to implement machine learning interventions in real-world settings and maximize the adoption of these interventions.

Introduction

Advances in predictive analytics and machine learning offer an opportunity to improve the diagnosis and management of acute conditions. A prominent use case for machine learning in health care is sepsis, a leading cause of death in US hospitals [1], which accounts for 1.7 million hospitalizations [2] and costs the US health system US $23 billion annually [3]. Machine learning

algorithms have been shown to outperform traditional screening scores in early sepsis detection [4]. Despite the rise in competing models and products for early sepsis detection [5], few machine learning models have been implemented as part of clinical care [6-8]. As a result, little evidence exists on the optimal integration of these sepsis models and other machine learning models into clinical workflows [9].

Early detection and treatment of sepsis is essential to decrease patient mortality [10]. Implementing standardized bundles can help ensure timely and proper sepsis care and has been associated with reductions in mortality [11-13]. Despite consensus that sepsis treatment bundles improve patient outcomes, only 49% of patients in US hospitals receive appropriate care [14]. The reasons for low compliance include the lack of a gold standard for sepsis diagnosis and the difficulty of rapidly mobilizing resources needed to treat individuals suspected of having sepsis [15,16].

Clinical decision support (CDS) systems may play a role in improving bundle compliance and delivery of timely treatment. CDS sepsis early warning systems leverage electronic health information to continuously stratify patients for the risk of sepsis and alert clinicians [17-22]. Unfortunately, many sepsis CDS systems fail to improve outcomes because of poor diagnostic accuracy and program implementation [23-25]. Eliciting perspectives directly from clinicians using such models can identify real-world barriers and facilitators impacting implementation efforts. Despite a growing body of literature on both physician perspectives of sepsis CDS implementation [26-28] and sepsis machine learning algorithms, more research is needed to understand clinician views on *black box* machine learning models that do not explain their predictions in a way that humans can understand [29,30]. To address this gap in the literature, this study identifies factors that affect the integration of a *black box* sepsis machine learning system into the workflows of frontline clinicians.

Methods

Setting

This study analyzes the implementation of the Sepsis Watch program at the Duke University Hospital (DUH). DUH is the flagship hospital of a multi-hospital academic health system with approximately 80,000 emergency department (ED) visits annually. According to our institutional definition for sepsis, over 20% of adults admitted through the DUH ED develop sepsis [31], and nearly 68% of sepsis occurs within the first 24 hours of hospital encounter [32].

Program Description

In a previous study, we designed a digital phenotype for sepsis using clinical data available in real time during the patient's hospital encounter. We then developed a deep learning model to predict a patient's likelihood of meeting the sepsis phenotype within the subsequent 4 hours [33,34]. The model analyzed 42,000 inpatient encounters and 32 million data points. Model inputs included static features (eg, patient demographics, encounter information, and prehospital comorbidities) and dynamic features (eg, laboratory values, vital signs, and medication administrations). The model pulls data from the electronic health record (EHR) and is updated every hour to ensure real-time analysis of sepsis risk.

Concurrent with model development, an interdisciplinary team of clinicians, administrators, and data scientists designed a workflow to translate outputs from the model into clinical action (Figure 1). The team created a web application to display all patients presenting to the ED and their risk of sepsis. In the application, every patient was classified and presented by the model as meeting sepsis criteria (black card), high risk of sepsis (red card), medium risk (orange card), or low risk (yellow card). Rapid response team (RRT) nurses are the primary users of the Sepsis Watch application and remotely monitor all patients in the ED. For patients meeting sepsis criteria or at high risk of sepsis, an RRT nurse conducts a chart review and calls the ED attending physician to discuss the patients' care pathway. If the attending physician agrees that the patient is likely to have sepsis, the RRT nurse supports the patient care team to ensure that sepsis care bundle items are ordered and completed. After the call, the RRT nurse continues to monitor the completion of the bundle items and follows up as needed with the ED attending physician or ED nurse.

Figure 1. Sepsis Watch workflow. ED: emergency department; RRT: rapid response team.
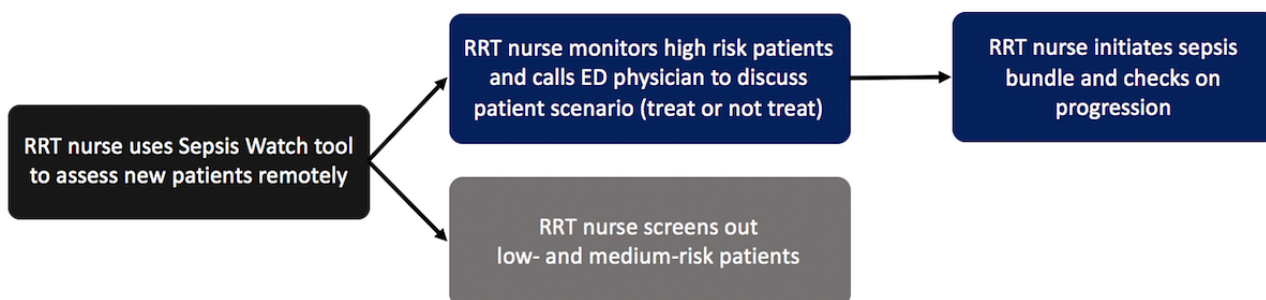
Before implementation, RRT nurses were extensively trained in person in the program workflow and application. ED physicians were informed about the program in faculty meetings and via email. Both nurses and physicians were educated on the model's aggregate performance measures relative to other methods, and visualizations of individual patient cases were

presented to demonstrate how the model could detect sepsis hours before the clinical diagnosis [30]. A full description of the planning and implementation process can be found elsewhere [32].

Study Design

A team of clinicians and social science researchers cocreated 2 interview guides: one for ED attending physicians and one for RRT nurses. The guide was designed to walk participants through each step of the workflow and probe for the associated barriers and facilitators. Subsequent questions in the interview guides covered training and dissemination of the new program, areas for improvement, and perceived utility. The guides were informed by the situational awareness model, which differentiates among 3 levels of situational awareness: (1) perception of relevant information, (2) comprehension of that information, and (3) anticipation of future events based on that information [35]. Although we drew from the situational awareness model to help support a high-level structure, we did not aim to examine the effect of Sepsis Watch on these levels or underlying mechanisms given our inductive and exploratory study approach. The interview guides were piloted among 3 clinicians to inform improvements to specific questions and overall structure.

ED leaders and RRT leaders invited physicians and RRT nurses to participate in semistructured interviews. As an operational project, participation was fully voluntary. From January 2019 to April 2019, we recruited 7 ED physicians (n=7; Table 1) and 8 RRT nurses (n=8; Table 2) to participate in the semistructured interviews. Although we used a convenience sampling approach to recruit participants, Tables 1 and 2 demonstrate how individual participants represent a diverse sample with regard to demographics, experience, and involvement in the design of the program. Of the 15 participants, 4 were involved in the design and development of Sepsis Watch.

Table 1. Characteristics of rapid response team nurse participants.

Participant	Experience as an RRT[a] nurse	Experience as a nurse	Was the participant involved in program development?
RRT nurse 1	7 months	4 years	No
RRT nurse 2	4 years	13 years	Yes
RRT nurse 3	5 years	10 years	No
RRT nurse 4	4 years	10 years	No
RRT nurse 5	3 years	5 years	No
RRT nurse 6	5 years	30 years	Yes
RRT nurse 7	3 years	4 years	Yes
RRT nurse 8	4 years	10 years	No

[a]RRT: rapid response team.

Table 2. Characteristics of emergency department attending participants.

Participant	Experience as attending physician at pilot site	Experience as attending physician	Was the participant involved in program development?
ED[a] attending 1	5 years	5 years	No
ED attending 2	2 years	2 years	No
ED attending 3	13 years	13 years	No
ED attending 4	5 years	5 years	Yes
ED attending 5	8 years	16 years	No
ED attending 6	9 months	9 months	No
ED attending 7	6 years	3 years	No

[a]ED: emergency department.

All interviews were conducted in person and face to face by the first author. Data collection started 4 months into the implementation of Sepsis Watch to give participants enough time to reflect on the initial rollout and describe any changes in workflow and perceptions. Following a semistructured interview format, participants were asked the same questions delineated in the interview guide, with flexibility for follow-up questions and probes. The interviews averaged 35 min in duration. Interviews were recorded with the written consent of the participants and transcribed verbatim. In the qualitative analytic process, data collection was terminated when no new themes or insights emerged (ie, thematic saturation) across both participant groups [36].

To analyze the transcripts, we followed a modified grounded theory approach, a widely used and established analytic method in social science research [37]. Grounded theory provides a systematic approach to derive and classify themes from qualitative data, such as interview transcripts. This approach

employs a coding process to organize qualitative data, in which text is labeled with *codes* or short phrases that reflect the meaning of sentences or paragraphs [38]. These codes are then used to generate higher-level themes that emerge as the major study findings.

In our study, coding was conducted in 3 phases. In the first phase, we used line-by-line coding to create tentative open codes closely grounded in the raw data. In the second phase, focused coding was employed to create higher-level categories and subcategories from the open codes. In the third phase, we selectively defined relationships among various categories. The final codebook was discussed and reviewed with members of the research team. Data were analyzed using NVivo Qualitative Data Analysis Software (version 12, QSR International). We queried codes to identify the most prevalent themes among the 2 participant groups. This study was approved by the Duke University Health System Institutional Review Board (Protocol ID: Pro00093721).

Results

A variety of themes emerged as important factors that shaped the integration of Sepsis Watch into routine clinical care (Table 3). Factors were grouped into 3 thematic areas: (1) perception of utility and trust, (2) implementation of Sepsis Watch processes, and (3) workforce considerations. For each area, we describe the corresponding subthemes with representative quotations.

Table 3. Thematic area and corresponding subthemes.

Thematic area	Subtheme
Perception of utility and trust	• Trust and accuracy • Perception of machine learning • Context-specific utility
Implementation of the Sepsis Watch program	• Tool layout and design • Value of human communication • Nurse strategies • Information flow challenges • Gaps in knowledge and understanding
Workforce considerations	• A new role—Sepsis Watch nurse • Skills and capabilities required for success

Perception of Utility and Trust

This area focuses on themes related to clinicians' attitudes toward the Sepsis Watch program, including trust and accuracy, broader perceptions about the role of machine learning in clinical practice, and the settings in which the program was perceived to be the most useful. Representative quotations are presented in Table 4.

Table 4. Representative quotations on perceived utility and trust.

Subtheme	Quote
Trust and accuracy	• "Sepsis Watch is very good at predicting patients and identifying patients who are septic and...we've had a lot of patients here that have actually come to our [CICU] unit from the ED[a] who have popped up on Sepsis Watch." [RRT[b] nurse] • "Blood cultures seem to weigh very heavily in the algorithm...I can pretty much bet you money that every single time I order blood cultures on somebody, sixty minutes later I'll get a phone call from Sepsis Watch that says they tripped positive...it means I was thinking about infection but I wasn't worried enough to pursue the true sepsis bundle." [ED physician] • "I had at least two patients who went to the ICU that I never got a Sepsis Watch call for, at all. So, I don't know how those got missed...The rest of them, so a lot of the false positives were like...COPD exacerbation or something like that." [ED physician] • "The initiative...just creates a lot more vigilance...I almost feel like I'm very cognizant of sepsis and almost like, imagining the Sepsis Watch people upstairs like, looking down on me...I'm honestly like, just waiting for their call. Like, can you imagine like, I was like, oh this must be them. So, in some ways I think that's good, that it has fostered vigilance." [ED physician]
Perception of machine learning	• "I think a big part of people not understanding [Sepsis Watch], including actually the ED doc, is if vitals are stable. We're not gonna treat because they look stable. I know but we're trying to catch it before it's unstable. And that's the biggest piece people don't get...fact that it's predictive like, hammering that in will help people see…we're trying to prevent the decline." [RRT nurse] • "Most people don't know much about [machine learning] and there's always this idea of like, you can't replace me and my training and that I'm standing in front of the patient telling you if they're septic or not." [ED physician]
Context-specific utility	• "It's probably a way more useful tool, not in the ED. In the ED, all we think about all the time is sepsis cause it's such a big part of our practice. So, that's why I think it doesn't apply well to us, but it would apply well in other settings where they don't think about or see or miss the bundle more often." [ED physician]

[a]ED: emergency department.

[b]RRT: rapid response team.

Trust and Accuracy

RRT nurses, the primary users of the Sepsis Watch tool, spoke positively about the accuracy of the Sepsis Watch model to detect sepsis. When asked how the Sepsis Watch tool faired against other CDS tools such as the National Early Warning Score that they had previously used [23], the RRT nurses overwhelmingly described the relative advantage of Sepsis Watch, both in diagnostic accuracy and data visualization. However, ED physicians described relatively less trust in the model. They thought certain components of the Sepsis Watch algorithm were too heavily weighted (eg, *blood cultures*). Physicians also noted that Sepsis Watch both missed some important sepsis cases and had false positives. To build more trust in the model, several physicians requested feedback on the success of Sepsis Watch and the specific cases that they missed but were detected by Sepsis Watch.

RRT nurses generally believed that the Sepsis Watch program improved sepsis detection and bundle compliance, especially because the implementation team provided them with data on the success of the program. ED physicians reported that Sepsis Watch had increased the vigilance and proactive culture about sepsis in the ED but were overall less positive than the RRT nurses. ED physicians described the difficulty in achieving the perfectly timed intervention in which the ED attending physician has had time to evaluate the patient but has still not diagnosed sepsis or completed the bundle items.

Perception of Machine Learning

Both RRT nurses and ED physicians said that they lacked the knowledge and understanding required to assess the validity of the machine learning model. Nurses reported feeling uncomfortable reviewing and assessing high-risk patients with minimal information and would often wait for more information to populate the medical record before having the confidence to call the ED physicians. Physicians also lacked knowledge about the model and the predictive nature of the model.

When asked about the role of machine learning in health care more broadly, physicians had varied responses. Some physicians noted a lack of knowledge, fear of overstepping, and resistance to change in medicine as potential barriers. Other physicians saw an opportunity to introduce machine learning to operations and logistics problems before CDS.

Context-Specific Utility

ED physicians felt that they were not necessarily the appropriate target adopters for Sepsis Watch. Respondents felt that ED attending physicians at large academic health centers were particularly adept at identifying and treating sepsis. Instead, they perceived Sepsis Watch to be most useful for residents who were still developing clinical skills, low-resource community settings, or hospitals with a poor track record for treating sepsis.

Implementation of Sepsis Watch Processes

This thematic area focuses on the implementation of the Sepsis Watch program, including the use of the tool and the interactions between RRT nurses and ED physicians. Although the nurses created their own communication strategies to facilitate interaction, barriers to positive interactions included challenges in information flow and gaps in understanding and knowledge. Representative quotations are presented in Table 5.

Table 5. Representative quotations on the implementation of Sepsis Watch processes.

Subtheme	Quote
Layout and design	• "It's just easy to navigate. You start at triage and go through the different tabs. The colors are easy...You quickly glance at it and you already have an idea of what you're getting yourself into...If you're used to navigating an iPhone, it's pretty easy to just figure it out." [RRT[a] nurse] • "I start to really go through the patient's chart and see what they presented to the emergency room for you know, what was their complaint, what's their past history, and then I'll look at their lab values and things like that and vitals and medications and stuff...The biggest thing I look for is the notes you know that the ED[b] staff are writing. You know, that kind of guides a lot." [RRT nurse]
Value of human communication	• "No matter how good the technology is, if the interface is bad no one's going to use it and then they're going to interpret that as the technology is bad...we use the RRT...like an air traffic controller in an airport that gets all this stuff, consolidates it and calls it out to the right people, until we figure out a way to do it through the computer interface." [ED physician]
Nurse strategies	• "This is how it goes. 'Hey, this is [person's name] from Sepsis Watch. How are you? Good. Okay, I'm calling about Mr Wallace in A-15. He's popped up at high risk for sepsis. I see that you know, he came in complaining of a cough. I see that you've already done like, a lactate, antibiotics. Are you thinking sepsis?' I try to put a piece of information to show I've done a chart review to show that this is not like, a cold call, that I've actually looked." [RRT nurse]
Informational flow challenges	• "If [the ED physicians] are busy with other patients, sometimes you cannot get communication with them on the first point of contact, so on your first phone call, they may be running a code in the resuscitation bay...then you have to wait about like, an hour or two to kind of get in touch with them." [RRT nurse] • "It's an interruption. I mean it's a random call at a random time that's completely disruptive to workflow. Every single call we get is completely disruptive to workflow. And when it's not giving me any new information, it's even less helpful." [ED physician] • "It would be hard to escalate to the ED physicians because we don't work with them, we're not there, we don't have that relationship with them. They don't know who we are, they don't really know what we do, so I think for me, then to be saying I feel like you need to start this patient on antibiotics...that wouldn't go down too well...If you were like, down, physically down in the ED with them, I think that would be a different case scenario." [RRT nurse] • "Part of the problem is, ED is such a team-based approach that it's often that you're so busy that I'm sort of doing my round around the ED caring for people and the resident's doing theirs, and the nurses doing theirs that you might not overlap frequently enough or adequately enough to convey that information to the people that need to know. For me to have to track them both down to give them that information would be burdensome and that's what would get in the way of flow in the ED." [ED physician]
Gaps in knowledge and understanding	• "In the beginning it was very difficult making those phone calls because I don't think that they understood exactly what Sepsis Watch was. There was a lot of like, 'who are you, what are you doing, is this is lawsuit type of thing?' They were worried...that if they decided not to treat...and then it turned into sepsis, that they were worried about potentially getting sued for malpractice." [RRT nurse] • "At first there was a little bit of unwieldiness with the actual bucket that we could sort the patients into, so in other words, what does it mean to place them into the [sepsis bundle] protocol, continue to watch them, or to say no, the source is not septic, they don't need to be watched any longer. But I think as time's gone on now, we're more comfortable with the different answers that they're looking for and that Sepsis Watch nurses are more comfortable guiding us to an answer." [ED physician] • "I think there's some areas for [ED physicians] to learn because a few questions I will get are like 'why does [Sepsis Watch] say they're high-risk because they don't look septic here.' Obviously, I don't know exactly why the app is populating them that way, so I think if they understood that we don't have all the bits of information that are making them a red card or a black card or yellow or orange...I just have how the computer model populates them into which color and I'm kind of going from there." [RRT nurse]

[a]RRT: rapid response team.

[b]ED: emergency department.

Layout and Design

RRT nurses frequently complimented the Sepsis Watch application for being easy to use and well designed. They described the benefits of visually delineating sepsis risk into colors (eg, red cards as high risk, orange cards as medium risk) and tracking patients across distinct tabs (eg, patients to be triaged, screened out for sepsis, and those in the sepsis bundle). Although RRT nurses use the Sepsis Watch dashboard to monitor important sepsis signs and symptoms (eg, lactate and white blood count), the EHR remains a valuable source for additional information. Some RRT nurses felt more comfortable

with the patient's chart and reported aggregating information presented in both systems.

Value of Human Communication

Both ED physicians and RRT nurses described the benefits of having RRT nurses as the effector arm of Sepsis Watch, especially in comparison with the more traditional *best practice alerts* (BPAs) through the EHR. Physicians described how BPAs frequently slow them down and that they are more likely to ignore the BPAs. In contrast, physicians reported that workflow interruptions from human interaction cause less alarm fatigue and get their attention immediately.

Nurse Strategies

To facilitate conversations with physicians, RRT nurses developed their own communication and workflow strategies. For example, rather than calling the ED physician for each patient with sepsis or for a high-risk case, nurses often grouped multiple patients together by area of the ED to minimize the number of calls. Similarly, RRT nurses avoided calling ED physicians before shift changes.

During the phone call itself, some RRT nurses ask "how are you" or "is this a good time to call" to gauge the physician's busyness before they continue the conversation. Other nurses presented information from their chart review to demonstrate that they had a working knowledge of the patient's case. Many RRT nurses cited the importance of being succinct, direct, and polite to maximize the chances of a positive interaction.

Informational Flow Challenges

Although most RRT nurses did not face many barriers in reaching the ED physicians on the phone, some described challenges. Nurses often described how the busy workflows of the ED physicians, coupled with the remote monitoring nature of the RRT, could impede information flow. For example, 1 nurse described challenges in calling physicians amid resuscitation efforts. Nurses hypothesized that communicating with ED physicians might be easier in person than via phone. Furthermore, nurses reported that a lack of working relationships between the ED physicians and RRT nurses before Sepsis Watch made it challenging to build rapport and communicate freely.

Physician respondents noted that calls, although brief, were still interruptions in their busy workflows, which decreased their receptivity for calls from the RRT nurse. In addition, physicians felt that targeting calls about sepsis detection and treatment to the ED attending physicians only required them to disseminate the information to the entire ED care team, such as residents and nurses.

Gaps in Knowledge and Understanding

RRT physicians reported that at the start of the program, the ED physicians were unfamiliar with the purpose of the program, the role of the RRT nurse, and the flow of the call. This initial unfamiliarity might have resulted in confusion and misunderstanding. For example, RRT nurses heard from some physicians that they feared being increasingly liable and accountable for proper sepsis treatment with Sepsis Watch rollout. Over time, collaboration between RRT nurses and ED physicians improved.

Even though the Sepsis Watch tool does not inherently provide explanations for risk scores, RRT nurses were sometimes asked to explain the risk score. This created a mismatch between what ED physicians and RRT nurses understood about the technology.

Workforce Considerations

This thematic area describes the workforce implications of the Sepsis Watch program. The program required the creation of a new professional *Sepsis Watch Nurse* role to translate the machine learning algorithm to the patient's bedside. It is also important to identify the skills and capabilities needed for nurses to successfully perform the duties of the new role. Representative quotations are presented in Table 6.

Table 6. Representative quotations on workforce implications.

Subtheme	Quote
A new role — Sepsis Watch nurse	• "It's been enlightening. You are Sepsis Watch nurse. You are watching sepsis you know, in the ED[a] and it's cool you know, it's a totally new job title under the RRT[b] role and a new responsibility and one I welcome. I think it's really good and I think having a nurse with good clinical judgement, hopefully, as being that second check." [RRT nurse] • "I would rather be looking at this than be walking around the unit doing turns, pulling up, boosting, cleaning, and putting out fires on the unit. So, this workflow has been nice like, it allows me to step back and use my mind in a different way." [RRT nurse] • "We're not here to contradict what they're already doing. If they tell me that they're not worried about sepsis, I don't disagree with them...I don't try to argue with them. They are the physician. They're the ones that know the patient. I'm looking at a computer screen. I don't actually see the patients themselves." [RRT nurse]
Skills and capabilities required for success	• "I think if you have a good clinical background and are familiar with sepsis and you're kind of familiar with how to treat sepsis and stuff that you can probably perform sepsis watch. I don't know that you necessarily have to be an RRT nurse...Sepsis Watch is so specific, if you've got a good gen[eral] med[icine] background, I think you could probably serve as a good sepsis watch nurse." [RRT nurse] • "I think getting people who only want to do it would be helpful. I think you'll find enough people who would want to do it I think making it not mandatory for people who don't want to do it. Recruit some people who do. Management support and buy in and hey, this is your job, it's important. And positive feedback as far as results, statistics." [RRT nurse]

[a]ED: emergency department.

[b]RRT: rapid response team.

XSL•FO

RenderX

A New Role: Sepsis Watch Nurse

RRT nurses took pride in their new role of a *Sepsis Watch Nurse*, especially given their participation in the program design and pilot implementation. More specifically, RRT nurses enjoyed the investigative and diagnostic role of the Sepsis Watch role. Although RRT nurses are empowered with information through Sepsis Watch, they recognized the boundaries of their own scope of practice and the need to continue to respect the professional autonomy of physicians.

Skills and Capabilities Required for Success

When asked about the skills and knowledge needed to be a good Sepsis Watch nurse, the RRT nurses mentioned good clinical judgment, knowledge of sepsis, and critical care experience. If nurses are unfamiliar with sepsis, they might rely too heavily on the model without using their own critical thinking skills. RRT nurses also explained the importance of strong communication skills to confidently speak with attending physicians whom they may not personally know. Although the Sepsis Watch nurse has to interact with a web-based dashboard, RRT nurses thought that strong computer skills were not necessary for the role, given the simplicity of the app. RRT nurses also recommended recruiting nurses interested in the role and the need to create buy-in through continuous feedback.

Discussion

In our study, we conducted interviews with ED physicians and RRT nurses to understand the factors affecting the integration of a machine learning tool into clinical workflows. We found 3 main thematic areas: (1) perception of utility and trust, (2) implementation of Sepsis Watch processes, and (3) workforce considerations, with 10 corresponding subthemes. Taken together, our findings show how RRT nurses can effectively monitor the outputs of a machine learning model and communicate their assessment to ED physicians. To our knowledge, this is the first qualitative research study to investigate the real-world implementation of a machine learning sepsis early warning system in practice.

RRT nurses had positive impressions of the layout and design of the Sepsis Watch tool. This may be partially explained by the participatory approach with which the Sepsis Watch solution was built. Clinicians provided frequent input from the design of the tool to its broader use in clinical workflows. Clinician preferences were incorporated to optimize the ease of use and utility for end users [39]. For example, the visual display of the risk of sepsis was simplified from a continuous risk scalar value into 3 brightly colored categories of risk (low, medium, and high) to reduce cognitive burden [26]. Furthermore, the simplicity of the tool allowed RRT nurses to integrate Sepsis Watch into their current clinical workflow instead of replacing workflows. The RRT nurses described how they still used the EHR and their own clinical judgment skills to contextualize the model outputs.

We also found that both RRT nurses and ED physicians had very limited prior exposure to machine learning–based CDS systems. The lack of machine learning foundational knowledge and firsthand experience made it more difficult for clinicians to trust the Sepsis Watch algorithm. For example, clinicians often felt uncomfortable trusting the Sepsis Watch prediction when they could not *see* clear signs and symptoms in their patients. Some physicians reported the need to know why and how the model predicted the outcome. Similarly, 2 previous studies examining predictive alerts for sepsis suggested that the perceived utility decreased when the model frequently identified clinically stable patients [27,28]. They also found that false positives from other nonsepsis etiologies could increase alarm fatigue. Thus, product developers must consider how limited model explainability and false positives threaten clinicians' trust in model outputs, particularly for patients without visible clinical symptoms. At the same time, the goal should be to optimize rather than maximize trust, in which clinicians maintain some skepticism of a tool's capabilities to prevent overreliance [40]. For example, we learned that even when physicians did not trust a model output, they still reported paying closer attention to a patient's clinical progression over time or ordering tests more quickly.

Despite these challenges, we also found that positive experiences with the tool and human connections improved clinician acceptance. RRT nurses described that their trust in the model increased from their personal experiences as the algorithm successfully predicted patients with sepsis. Physicians suggested that receiving feedback on patients with sepsis who they had personally missed diagnosing but who were correctly identified by Sepsis Watch would build trust in the model. Future implementation efforts may incorporate feedback loops to improve clinician adoption of machine learning models. As machine learning products become more widespread, health professional schools should incorporate foundational machine learning courses into their curriculum to build baseline literacy [41]. Health care organizations should provide training and educational resources and conferences for their existing clinical staff [42]. Model developers should develop clear product labels to help clinicians understand when and how to appropriately incorporate machine learning model outputs into clinical decisions [43].

As health systems start to implement *black box* models as part of routine care, this study shows the feasibility of leveraging a small team of nurses to communicate machine learning outputs to a larger cohort of ED physicians. Previous attempts at sending automated CDS alerts directly to the treating provider have been associated with high levels of alert fatigue [23]. In our program, RRT nurses mitigate alarm fatigue by screening the patients first, holding the providers more accountable for meeting bundle requirements, and by adding a human connection to the model output. Adding a human intermediary has its own challenges, such as interrupting busy ED workflows and additional delays with remote monitoring. For example, despite the model's high predictive value within the first hour of ED presentation [44], nurses sometimes waited for more clinical information to populate in the medical record before calling physicians about patients flagged as being at high-risk for sepsis. In future iterations of the program, these human-made delays need to be anticipated and addressed through training or workflow design to ensure patient safety.

Future programs that deploy clinicians as an *effector arm* for model outputs must also consider how to best recruit, train, and deploy their machine learning–ready workforce. We found that creating a new, specialized role for the program allows nurses enough time and patient volume to build unique expertise and effective strategies to communicate with physicians. Recruiting clinicians with strong interest in the program, familiarity with sepsis, and strong communication skills is critical.

We were able to use the findings from these interviews to drive program improvements and to inform the scale. For example, to improve the user interface of the tool, we allowed for more space for free text comments from nurses, and we de-emphasized visual displays of model *trend lines* for sepsis risk. To streamline the workflow and reduce the burden on ED physicians, we no longer required nurses to call ED physicians if they had already clearly ruled out sepsis or started treatment. We also identified broader strategies to build trust and accountability in machine learning tools, such as educating clinicians on model performance and utility within the local context, respecting professional discretion, and engaging end users early and often throughout program design and implementation [30].

Limitations

This study has several limitations. First, its findings need to be generalized with caution to other settings or applications that differ in organizational structure, capacity, and professional norms and practices. Other settings that also use a participatory approach to intervention development may result in unique tool layouts or workflows. Second, the study largely focuses on workflow integration and does not explore the multi-year planning and stakeholder engagement process crucial for the successful launch of Sepsis Watch. Similarly, only frontline clinicians were interviewed despite the large number of stakeholders involved in project development and maintenance (eg, organizational leadership, hospital administrators, data scientists) or impacted directly by the program (ie, patients; [45,46]). However, for frontline clinicians, having 2 distinct respondent groups in the study allowed for triangulation and strengthening of the analysis. Finally, although we included clinicians involved in the development of the Sepsis Watch program in our sample, given their unique expertise and insight, their participation could have biased the findings to frame Sepsis Watch more positively. Thus, further studies with a larger sample size may use survey approaches to quantify factors influencing the adoption of machine learning CDS tools and examine variations by clinician characteristics, including involvement in program development. Such approaches should consider using well-studied theoretical models for instrument development, such as the Unified Theory of Acceptance and Use of Technology [47].

Conclusions

Although previous studies have studied factors affecting the implementation of CDS tools [48], the use of black box models in health care settings presents unique challenges and opportunities related to trust and transparency. Previous studies exploring the implementation of these models have focused on surveying clinician perspectives on *future* rollout [42,49-51]. Unfortunately, these studies do not uncover providers' real-life experiences using artificial intelligence tools in current practice. More research is needed to understand the real-world barriers and facilitators to the design and implementation of machine learning products. Understanding how these factors interact in diverse contexts can inform implementation strategies to ensure adoption. Although we used our findings to inform program improvements locally, our learning can help other health organizations that are planning to integrate machine learning tools into routine practice.

Conflicts of Interest

MS, WR, AB, NB, and CO are named inventors of the Sepsis Watch deep learning model, which was licensed from Duke University by Cohere Med Inc. These authors do not hold any equity in Cohere Med Inc. No other authors have relevant financial disclosures.

References

1. Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. JAMA Netw Open 2019 Feb 1;2(2):e187571 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7571] [Medline: 30768188]
2. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. J Am Med Assoc 2017 Oct 3;318(13):1241-1249 [FREE Full text] [doi: 10.1001/jama.2017.13836] [Medline: 28903154]

XSL•FO
RenderX

3. Torio CM, Moore BJ. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204. Healthcare Cost and Utilization Project (HCUP). URL: https://europepmc.org/abstract/med/27359025 [accessed 2020-11-13]

4. Islam MM, Nasrin T, Walther BA, Wu C, Yang H, Li Y. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019 Mar;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]

5. Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020 Mar;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]

6. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. Crit Care Med 2019 Nov;47(11):1485-1492. [doi: 10.1097/CCM.0000000000003891] [Medline: 31389839]

7. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017;6(2):e000158 [FREE Full text] [doi: 10.1136/bmjoq-2017-000158] [Medline: 29450295]

8. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res 2017;4(1):e000234 [FREE Full text] [doi: 10.1136/bmjresp-2017-000234] [Medline: 29435343]

9. Sendak D, Arcy J, Kashyap S, Gao M, Corey K, Ratliff B, et al. A path for translation of machine learning products into healthcare delivery. EMJ Innov 2020 Jan 27:- epub ahead of print. [doi: 10.33590/emjinnov/19-00172]

10. Nguyen HB, Corbett SW, Steele R, Banta J, Clark RT, Hayes SR, et al. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. Crit Care Med 2007 Apr;35(4):1105-1112. [doi: 10.1097/01.CCM.0000259463.33848.3D] [Medline: 17334251]

11. Ferrer R, Artigas A, Levy MM, Blanco J, González-Díaz G, Garnacho-Montero J, Edusepsis Study Group. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. J Am Med Assoc 2008 May 21;299(19):2294-2303. [doi: 10.1001/jama.299.19.2294] [Medline: 18492971]

12. Levy MM, Evans LE, Rhodes A. The surviving sepsis campaign bundle: 2018 update. Crit Care Med 2018 Jun;46(6):997-1000. [doi: 10.1097/CCM.0000000000003119] [Medline: 29767636]

13. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to treatment and mortality during mandated emergency care for Sepsis. N Engl J Med 2017 Jun 8;376(23):2235-2244 [FREE Full text] [doi: 10.1056/NEJMoa1703058] [Medline: 28528569]

14. Barbash IJ, Davis B, Kahn JM. National performance on the medicare SEP-1 Sepsis quality measure. Crit Care Med 2019 Aug;47(8):1026-1032 [FREE Full text] [doi: 10.1097/CCM.0000000000003613] [Medline: 30585827]

15. Kissoon N. Sepsis guideline implementation: benefits, pitfalls and possible solutions. Crit Care 2014 Mar 18;18(2):207 [FREE Full text] [doi: 10.1186/cc13774] [Medline: 25028802]

16. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human Septic shock. Crit Care Med 2006 Jun;34(6):1589-1596. [doi: 10.1097/01.CCM.0000217961.75225.E9] [Medline: 16625125]

17. Arabi YM, Al-Dorzi HM, Alamry A, Hijazi R, Alsolamy S, Al Salamah M, et al. The impact of a multifaceted intervention including sepsis electronic alert system and sepsis response team on the outcomes of patients with sepsis and septic shock. Ann Intensive Care 2017 Dec;7(1):57 [FREE Full text] [doi: 10.1186/s13613-017-0280-7] [Medline: 28560683]

18. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. J Am Med Inform Assoc 2017 Jan;24(1):88-95. [doi: 10.1093/jamia/ocw056] [Medline: 27225197]

19. Nelson JL, Smith BL, Jared JD, Younger JG. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. Ann Emerg Med 2011 May;57(5):500-504. [doi: 10.1016/j.annemergmed.2010.12.008] [Medline: 21227543]

20. Parshuram CS, Dryden-Palmer K, Farrell C, Gottesman R, Gray M, Hutchison JS, Canadian Critical Care Trials Groupthe EPOCH Investigators. Effect of a pediatric early warning system on all-cause mortality in hospitalized pediatric patients: the EPOCH randomized clinical trial. J Am Med Assoc 2018 Mar 13;319(10):1002-1012 [FREE Full text] [doi: 10.1001/jama.2018.0948] [Medline: 29486493]

21. Sawyer AM, Deal EN, Labelle AJ, Witt C, Thiel SW, Heard K, et al. Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. Crit Care Med 2011 Mar;39(3):469-473. [doi: 10.1097/CCM.0b013e318205df85] [Medline: 21169824]

22. Umscheid CA, Betesh J, van Zandbergen C, Hanish A, Tait G, Mikkelsen ME, et al. Development, implementation, and impact of an automated early warning and response system for sepsis. J Hosp Med 2015 Jan;10(1):26-31 [FREE Full text] [doi: 10.1002/jhm.2259] [Medline: 25263548]

23. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. Crit Care Med 2019 Jan;47(1):49-55 [FREE Full text] [doi: 10.1097/CCM.0000000000003439] [Medline: 30247239]

XSL•FO
RenderX

24. Harrison AM, Gajic O, Pickering BW, Herasevich V. Development and implementation of sepsis alert systems. Clin Chest Med 2016 Jun;37(2):219-229 [FREE Full text] [doi: 10.1016/j.ccm.2016.01.004] [Medline: 27229639]

25. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Qual Saf 2019 Sep;28(9):762-768 [FREE Full text] [doi: 10.1136/bmjqs-2018-008765] [Medline: 30872387]

26. de Vries A, Draaisma JM, Fuijkschot J. Clinician perceptions of an early warning system on patient safety. Hosp Pediatr 2017 Oct;7(10):579-586 [FREE Full text] [doi: 10.1542/hpeds.2016-0138] [Medline: 28928156]

27. Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. Crit Care Med 2019 Nov;47(11):1477-1484. [doi: 10.1097/CCM.0000000000003803] [Medline: 31135500]

28. Guidi JL, Clark K, Upton MT, Faust H, Umscheid CA, Lane-Fall MB, et al. Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. Ann Am Thorac Soc 2015 Oct;12(10):1514-1519. [doi: 10.1513/AnnalsATS.201503-129OC] [Medline: 26288388]

29. Elish MC. The Stakes of Uncertainty: Developing and Integrating Machine Learning in Clinical Care. In: Ethnographic Praxis in Industry Conference Proceedings. 2019 Feb 1 Presented at: EPIC'19; October 11, 2018; Honolulu, Hawaii. [doi: 10.1111/1559-8918.2018.01213]

30. Sendak M, Elish M, Gao M, Futoma J, Ratliff W, Nichols M, et al. The Human Body is a Black Box: Supporting Clinical Decision-making With Deep Learning. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020 Presented at: FAT'20; January 3-9, 2020; Barcelona, Spain. [doi: 10.1145/3351095.3372827]

31. Lin A, Sendak M, Bedoya A, Clement M, Brajer N, Futoma J, et al. Evaluating sepsis definitions for clinical decision support against a definition for epidemiological disease surveillance. bioRxiv. Preprint posted online May 31, 2019. URL: https://www.biorxiv.org/content/10.1101/648907v1 [accessed 2020-11-11]

32. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020 Jul 15;8(7):e15182 [FREE Full text] [doi: 10.2196/15182] [Medline: 32673244]

33. Futoma J, Hariharan S, Heller K. Learning to detect sepsis with a multitask Gaussian process RNN classifier. arXivstat Preprint posted online June 13, 2017 [FREE Full text]

34. Futoma J, Hariharan S, Sendak M, Brajer N, Clement M, Bedoya A, et al. An Improved Multi-Output Gaussian Process RNN with Real-Time Validation for Early Sepsis Detection. arXivstat. 2017. URL: http://arxiv.org/abs/1708.05894 [accessed 2019-04-14]

35. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 2016 Nov 23;37(1):32-64. [doi: 10.1518/001872095779049543]

36. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]

37. Charmaz K. Constructing Grounded Theory. Thousand Oaks, CA: Califage Publications; 2006.

38. Creswell J, Creswell J. Research design: Qualitative, Quantitative, and Mixed Methods Approaches. Los Angeles, USA: SAGE Publications; 2018.

39. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

40. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]

41. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med 2018;1:54 [FREE Full text] [doi: 10.1038/s41746-018-0061-1] [Medline: 31304333]

42. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med 2019;2:28 [FREE Full text] [doi: 10.1038/s41746-019-0106-0] [Medline: 31304375]

43. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. NPJ Digit Med 2020;3:41 [FREE Full text] [doi: 10.1038/s41746-020-0253-3] [Medline: 32219182]

44. Bedoya A, Futoma J, Clement M, Corey K, Brajer N, Lin A, et al. Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open 2020 Jul;3(2):252-260 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa006] [Medline: 32734166]

45. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClinicalMedicine 2019 Mar;9:52-59 [FREE Full text] [doi: 10.1016/j.eclinm.2019.03.001] [Medline: 31143882]

46. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018 Mar 12;8(1):4330 [FREE Full text] [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]

47.   Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]

48.   van de Velde S, Heselmans A, Delvaux N, Brandt L, Marco-Ruiz L, Spitaels D, et al. A systematic review of trials evaluating success factors of interventions with computerised clinical decision support. Implement Sci 2018 Aug 20;13(1):114 [FREE Full text] [doi: 10.1186/s13012-018-0790-1] [Medline: 30126421]

49.   Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. Artif Intell Med 2020 Jan;102:101753. [doi: 10.1016/j.artmed.2019.101753] [Medline: 31980092]

50.   Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019 Mar 20;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]

51.   Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 9;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]

Abbreviations

BPA: best practice alert
CDS: clinical decision support
DUH: Duke University Hospital
ED: emergency department
EHR: electronic health record
RRT: rapid response team

Original Paper

Usability and Acceptability of a Mobile App to Help Emerging Adults Address their Friends' Substance Use (Harbor): Quantitative Study

Kyle Michael Bennett[1], PhD, MSW; Kelly Lynn Clary[2], PhD, MSW; Douglas Cary Smith[1], PhD, LCSW; Carol Ann Lee[1], MSW

[1]University of Illinois Urbana-Champaign, Urbana, IL, United States
[2]Texas State University, San Marcos, TX, United States

Corresponding Author:
Kyle Michael Bennett, PhD, MSW
University of Illinois Urbana-Champaign
1010 West Nevada Street
Urbana, IL, 61801
United States
Phone: 1 2177660755
Email: kmbennet@illinois.edu

Abstract

Background: Technology-assisted intervention and prevention strategies present opportunities for substance use–related research with emerging adults (EAs) and their peers. Emerging adulthood is a developmentally distinct period in which individuals between the ages of 18 and 29 years undergo unique emotional, cultural, developmental, and biological changes as they transition into adulthood. Crowdsourcing, or gathering feedback from a large group within web-based communities, offers researchers a unique and cost-effective way to obtain large amounts of information in a short period.

Objective: This paper presents market feedback obtained via Amazon's Mechanical Turk from EAs (N=458) on the acceptability and utility of brief intervention scripts for a smartphone app currently under development. The mobile app, Harbor, teaches friends of EAs with substance use problems effective and supportive strategies for helping their friend make changes in their substance use behavior.

Methods: We examined feedback on the wording of the intervention scripts and estimated the market size of EAs who may use this app. Furthermore, we calculated correlations between script ratings and measures of personal risky drinking (ie, Alcohol Use Disorder Identification Test) and the participants' use of confrontational, enabling, or supportive behaviors with an existing friend.

Results: Approximately half of our sample (208/458, 45.4%) indicated that they had a close friend for whom they had concerns about their substance use, suggesting a potentially high demand for an app such as Harbor. Initial findings suggest that peers who engage in less enabling behaviors with friends who have a substance use problem exhibited lower risky drinking behaviors overall ($r_{206}=-0.501$; $P<.001$). Concerning acceptability, 98.0% (449/458) of the sample rated the scripts' dialogue as either somewhat, moderately, or extremely realistic (mean 3.92, SD 0.48) on 5-point Likert scale items. Finally, 95.4% (437/454) of respondents indicated that the scripts would be at least slightly helpful for training peers to help their friends with substance use issues. Finally, individuals who were better able to identify enabling language in enabling scripts self-reported fewer enabling behaviors toward their friend's substance use ($r_{206}=-0.236$; $P=.001$).

Conclusions: There exists a demonstrated level of desirability and acceptability among EAs for a mobile app such as Harbor. EAs who wish to engage in more supportive behaviors with their friends who engage in substance use and who are amenable to assisting their friends with sobriety likely would use and benefit from this app.

(J Med Internet Res 2020;22(11):e16632) doi:10.2196/16632

KEYWORDS

XSL•FO
RenderX

Introduction

Background

National surveys show that emerging adults (EAs; 18-29 years) in the United States have higher rates of substance use than any other age group [1]. To make matters worse, EAs also have inferior treatment engagement rates compared with other age groups [1]. In a nationally representative sample of EAs (N=19,312), only 11% (n=2124) of those diagnosed as having a substance use disorder received related treatment [2]. Given this treatment gap, it is important to think about concerned significant others (CSOs), such as peers, of EA substance users as potential first responders. The reach of social networks in the United States peaks during the emerging adulthood years, which points toward the importance of close friends and the influence they can have during this time [3]. This paper presents data from a crowdsourcing study to inform the development of a mobile app that harnesses the social support of such peers for EA substance users.

Online survey research (ie, crowdsourcing) allows investigators a flexible and convenient means of collecting data on a wide variety of issues and topics. Crowdsourcing is the practice of collecting information or feedback on a project or task by enlisting the assistance of large samples of people, typically via the internet. In addition to the relative ease with which researchers can collect online data, especially with EAs, crowdsourcing is as reliable as or even more reliable than traditional EA sampling methods [4]. Crowdsourcing respondents tend to be more ethnically diverse [4], which is particularly important in the development of substance use interventions. That is, individuals in clinical trials are less ethnically diverse, less clinically impaired, more highly educated, and more frequently employed than individuals receiving treatment in community settings [5]. In short, online crowdsourcing may assist intervention developers in acquiring more diverse samples of EAs.

Following the principles of community-based participatory research (CBPR) in the development of the Harbor mobile app (eg, collaborating with community members and translating knowledge gained through the partnership into specific actions), we conducted a crowdsourcing study with a large sample of EAs (N=458). Combining aspects of CBPR with this crowdsourcing project may encourage a sense of empowerment and collective ownership in EA substance use research [6]. Participants gave feedback on whether they considered themselves to have a friend experiencing substance use issues and how they viewed the intervention scripts for Harbor. In addition, among those with friends with substance use problems, we analyzed associations between their intervention script feedback and measures of actual interactions they had with friends. Finally, we analyzed whether and how participants' personal alcohol use was associated with script ratings.

Harbor helps EAs (aged 18-29 years) assist their friends in making changes to their substance use behaviors. It is an adaptation of a family therapy model called the Community Reinforcement Approach With Family Training (CRAFT), where CSOs learn communication skills to help individuals who demonstrated risky or problematic substance use behaviors [7].

Objectives

Specifically, we sought to answer the following 3 questions:

1. What is the potential market size of EAs who may use the Harbor mobile app?
2. What do potential Harbor users think of our intervention screenshots/scripts?
3. How do screenshot/script ratings associate with self-reported participant behaviors in response to existing friends experiencing substance use problems?

Regarding the hypotheses, we first proposed that a large proportion of EAs would have friends with substance use problems. Second, based on prior acceptability studies for our face-to-face model, we hypothesized that participants would rate the scripts favorably. Third, we hypothesized that intervention script ratings would correlate negatively with actual self-reported interactions with their friends. Specifically, more confrontational or enabling friends would view the supportive intervention script (ie, CRAFT procedures script) as less helpful. Similarly, confrontational or enabling friends would rate the intervention scripts as less confrontational or enabling, respectively. We based hypotheses on the logic that such friends possess biases toward how they currently interact with their EA friends who misuse substances. Finally, we hypothesized that participants with riskier alcohol use would rate scripts as less enabling and self-report more frequent enabling behaviors toward their friends' substance use compared with individuals with less risky alcohol use.

Methods

CBPR

CBPR promotes participant involvement and feedback in all stages of the research process and has resulted in greater inclusion of users' experiences in behavioral health intervention development processes overall [8-11]. As part of this effort, there has been an increased focus on the expansion of patient-centered medicine [12], shared decision-making practices [13], and patient-friendly forms of communicative implementation as a whole [14]. Central to the CBPR ideology is the notion that by embracing and incorporating the experiences of end users, researchers reduce the risk of misunderstanding individuals' health-seeking behaviors, which ultimately lowers the risk of undermining clinical practice [15]. Cognizant of and responsive to unique cultural, social, and economic factors affecting EA communities, CBPR is a move toward reconciliation, reciprocity, and production of culturally relevant prevention measures for this population. In this regard, a CBPR approach is well suited for promoting the utilization of substance use treatment among EAs.

Integrating end user feedback during the initial development may help improve the effect sizes of technological interventions, which typically have small effect sizes [16,17]. Thus, lay knowledge and expertise may help improve intervention outcomes in the field [18]. In addition, sustaining intervention or prevention strategies with a target population proves difficult

XSL·FO
RenderX

[19], but seeking input from community members in research design and achieving shared health outcomes increases program sustainability [20]. This study asked EAs for initial feedback on script wording and the app's quality to drive revisions. Individuals with lived experience can later review these revisions.

Harbor: Background and Intervention Philosophy

In 2 prior studies, the authors adapted the CRAFT model for use with peer dyads [21,22]. At its core, CRAFT is a behavioral family therapy model for the CSOs of individuals with substance use disorders [7], and most research on CRAFT targets spouses. The CRAFT model emphasizes that CSOs attempt to interrupt their loved ones' substance use by (1) arranging prosocial activities at times when they would likely use, (2) refraining from enabling their substance use by not doing pleasant things with them while they are using or doing things to prevent the natural consequences of their use, and (3) avoiding confrontation while their significant other is intoxicated. Essentially, CRAFT trains CSOs to reward substance-free behaviors and not reward substance-using behaviors. The overarching idea is that individuals will choose a substance-free life if it is more rewarding than one involving substances.

Previous research indicates that EAs are amenable to helping their close friends by learning the behaviors associated with the CRAFT model. However, they were less likely to want to disassociate with friends when their friends were using [21]. Thus, the peer dyad version of CRAFT, called the Peer-Enhanced Community Reinforcement Approach (Peer-CRA), de-emphasized this feature, which appears more suitable for those in romantic and cohabitating relationships. In addition, in a preliminary study, both peers and EAs with substance use problems who received Peer-CRA significantly reduced their substance use [22]. Despite promising findings, one issue in a previous face-to-face study was that most peers only attended 1 session and did not receive the full CRAFT model as planned. Ideally, the development of a mobile app will aid in the delivery of this content, supporting peer first responders whose EA friends may have not yet received any sort of substance use intervention and/or treatment.

Harbor App Design and Features

The third author of this manuscript (DS) and a team of 4 EAs developed Harbor, which comprises 4 main modules. First, peers enter information about their EA friend's substance use to get feedback on the seriousness of their friend's problem. This module adapts the principles of brief motivational interventions by providing normative feedback, comparing an individual's use with well-established norms. Such interventions are efficacious at reducing substance use by correcting misperceived norms about substance use prevalence [23]. This novel feature of Harbor delivers normative feedback to peers with hopes of engaging them in helping processes by raising awareness about the gravity of their friend's substance use. This module may be useful for peers who are ambivalent about helping a friend with substance use problems. The second module teaches peers the principles of CRAFT through a series of mock text dialogues demonstrating confrontational, enabling, and supportive (ie, per the CRAFT intervention) responses to

their EA friend's retelling of a substance use–related incident (Multimedia Appendices 1-3). This study focused primarily on this feature of Harbor. Third, Harbor contains a module designed to discuss the idea of arranging competing prosocial activities at times when the EA friend would normally use substances. Finally, Harbor concludes with a module focused on training EA peers on how to have a discussion with their EA friends about seeking treatment.

The Market Need for Harbor

A recent meta-analysis found that *only about 20% (*n=10) of studies with EA samples from noncollege settings used technology-assisted interventions [24]. Furthermore, most technology-assisted interventions included were preventative in nature and mostly targeted alcohol use. The Harbor app addresses these limitations. First, it galvanizes EA peers to act as first responders for their close friends who demonstrate risky substance use behaviors. Second, it teaches EA peers' supportive skills to address multiple forms of substance use (ie, alcohol and cannabis use). In addition, a systematic review found that technology-assisted interventions worked equally well compared with the more traditional interventions for EAs in noncollege settings [24]. This finding suggests that the Harbor app may have expanded utility beyond traditional college settings where most research with EA samples occurs.

In examining the market for mobile apps related to Harbor, we found a dearth of programs designed to address the needs of concerned EAs with close friends who use substances. Most apps on the market target individuals attempting to ameliorate their own substance use issues and use common intervention strategies such as daily affirmations, drink tracking, and networking. Apps such as *Blood Alcohol Tracker*, *AlcoDroid*, *Substance Abuse and Addiction Assessments*, *Cravings Manager*, and *12 Steps AA Companion* currently exist for Apple, Android, and Blackberry users [25]. Each app has a unique self-reporting function and interfaces primarily with the client. Unlike the Harbor app, these apps do not involve another person to assist with changing behaviors. Although effective in reducing personal substance use or connecting individuals who may have substance use issues, these and other programs do not specifically focus on EAs. Furthermore, rather than focusing solely on recovery-oriented support (eg, Al-Anon Speaker Tapes and Al-Anon Audio Companion), a program such as Harbor could provide harm reduction alternatives to EA peers with concerns about a friend's substance use along the full severity spectrum.

Participant Recruitment

The authors obtained institutional review board approval before implementing the study procedures and informed consent from participants before their engagement with the app. Researchers conducted participant recruitment via Amazon's Mechanical Turk (MTurk) program from April 19, 2016, to September 2, 2016. MTurk is an online pool of workers who complete surveys and similar tasks for remuneration. In addition, surveys completed by MTurk workers have yielded reliable data [26], and online MTurk studies have successfully replicated findings originally drawn from in-person studies [27].

Despite the promise of MTurk for data collection, additional safeguards are necessary to enhance data quality. First, we used qualification screening to confirm participants' eligibility. After completing this qualification test and if eligible to participate in the study, participants completed the full survey. Participants were between the ages of 18 and 29 years and reported having consumed at least one alcoholic beverage in the past year. We eliminated survey responses completed too quickly (<5 min) and those missing validity check questions from the participant pool [27,28]. Finally, it is possible, although extremely unlikely, that one person would have multiple accounts and complete the full survey multiple times. However, Amazon's requirement of providing a valid social security number during account setup should prevent one person from registering multiple times.

Measures

Intervention Script Ratings

Participants read short text conversations used to teach the principles of enabling substance use, confronting another's substance use, or encouraging change in their friend's substance use via nonjudgmental responses that also encouraged substance-free activities (ie, supportive behaviors). Using 5-point Likert scales, participants then rated these scripts on how enabling (1=extremely discouraging use to 5=extremely enabling use), confrontational (1=extremely nonconfrontational to 5=extremely confrontational), or supportive (1=extremely unsupportive to 5=extremely supportive) they were. Furthermore, the authors prompted participants to respond with how accurately the scripts depict real-life conversations between EAs (1=extremely unrealistic to 5=extremely realistic) and how useful the scripts would be in teaching them how to support their friend's sobriety (1=not helpful at all and 5=extremely helpful). Multimedia Appendices 1-3 show examples of confrontational, enabling, and supportive scripts respectively, as well as script rating items for each. Finally, respondents rated how likely they would be to use this mobile phone app or others like it based on various referral sources (1=completely unlikely and 5=completely likely).

Self-Reported Interactions With Friends

The authors asked eligible participants, "Do you currently have any friends for whom you have concerns about their substance use?" Those who replied affirmatively completed the Significant Other Behavior Questionnaire (SBQ), a valid and reliable (α=.85) measure of the significant other's responses to an identified individual's substance use [29-31]. Authors amended the wording of individual items of the SBQ to achieve greater relevance for EAs, close friend dyads (eg, "friend" replaced "significant other"). In addition, the adapted version of the SBQ included 1 revised item to eliminate the assumption that all peers live together (eg, "Do you refuse to be around your friend when they are drinking or using?" versus "Do you refuse to be home when they are drinking?") and used "drinking and drug use" instead of "alcohol use" to account for multiple substances [20].

The amended SBQ contains 5-point Likert scale items (1=completely unlikely and 5=completely likely) clustered within 4 subscales: *support sobriety (SS)*, *support substance use (SSU; ie, enabling)*, *punish substance use (PSU; ie, confronting)*, and *withdrawal from friend (WFF)*. The SS subscale (8 items, range 8-40) measures the frequency with which a peer uses positive supportive strategies with their friend to encourage their friend to remain sober (eg, "Do you spend more time with your friend when they are not drinking or using drugs?"). The SSU subscale (6 items, range 6-30) measures the frequency with which a peer has encouraged their friend to use substances (eg, "Do you bring alcoholic beverages/drugs to your friend?"). The PSU subscale (5 items, range 5-25) measures the frequency with which a peer uses coercive or otherwise negative strategies to stop their friend from using substances (eg, "Do you have arguments regarding drinking or drug use when your friend is using?"). The WFF subscale (5 items, range 5-25) measures the frequency with which a peer attempts to withdraw from their friend while they are using substances (eg, "Do you refuse to be around your friend when they are drinking or getting high?").

Substance Use

To measure participants' personal substance use, we used the Alcohol Use Disorder Identification Test (AUDIT), a valid and reliable (α=.88) measure of risky alcohol use [32-34]. Along with the AUDIT, we measured peers' related substance use problems with the Substance Problem Scale (SPS, past month version), a 16-item scale from the larger Global Assessment of Individual Needs-I. Each item consists of a 4-point Likert scale indicating the last time an individual experienced a particular substance use–related issue (0=never, 1=more than a year ago, 2=past year, and 3=past month). We recoded responses to SPS items to reflect the past year problems categorically (0=not past year, 1=past year/month). The SPS contains items consistent with the Diagnostic and Statistical Manual of Mental Disorders criteria for substance use disorders and some additional items related to risky substance use behaviors (eg, hiding use) [35]. Furthermore, SPS has been used in various EA studies [36-38], demonstrated good reliability (α=.85) in predicting substance use disorders with a nationally representative sample of EAs [39], cross-validated with independently made psychiatric diagnoses [40], and is correlated with perceived EA status [38].

Data Analysis

The authors conducted analyses using SPSS 25.0 [41], first checking distributions for normality and assessing missing data, which were minimal. Nevertheless, analysts used multiple imputations through which statistical software imputed missing variables 5 times, analyzed each of the completed data sets, and pooled the results of each analysis into a final data set. The final results did not significantly differ from the original, unadjusted analyses. The authors present descriptive analyses of demographic data, individual SBQ items and SBQ subscales, and full-scale scores from the SPS and AUDIT. Furthermore, the authors estimated the potential market size of individuals that may benefit from using the mobile app. In addition, analysts calculated Pearson product-moment correlations between the SBQ, its subscales, and intervention script ratings and intercorrelations between SBQ subscales. Finally, analysts calculated Pearson correlations between these intervention script ratings and personal measures of risky drinking behaviors (ie,

AUDIT) as well as between the SBQ subscale and AUDIT scores.

Results

Demographics

Table 1 illustrates the demographic characteristics of our sample. On average, the full EA sample included participants who were 25.4 years old; among them, 58.1% (266/458) identified as male and 72.7% (333/458) identified as White. We note that the proportion of the sample working 35 or more hours per week (248/458, 54.1%) is lower than that reported in many national, multisite clinical trials [5]. Participants reported a median, past year income of roughly US $23,000, and a mean MacArthur Scale of Subjective Social Status [42] of 4.29. Our sample had a mean SPS (past year) score of 2.56 (range 0-16) and a mean AUDIT score of 6.38 (range 0-36). In both cases, higher scores represent greater issues pertaining to substance use.

Table 2 contains the SBQ individual item results as well as subscale scores.

Table 1. Participant demographic characteristics (N=458).

Characteristics	Value
Age (years), mean (SD)	25.39 (2.84)
Age at first substance use (years), mean (SD)	17.18 (2.60)
Male, n (%)	270 (59.0)
Race/ethnicity, n (%)	
White	333 (72.7)
Asian	47 (10.3)
African American	32 (7.0)
Hispanic/Chicano/Latino	31 (6.8)
American Indian/Alaskan Native	5 (1.1)
Native Hawaiian/Pacific Islander	1 (0.2)
Other	9 (2.0)
Education/occupation, n (%)	
Any postsecondary education	270 (59.0)
Employed full time	248 (54.1)
Employed part time	117 (25.5)
Unemployed	93 (20.3)
Unenrolled and unemployed	40 (8.7)
Current enrollment status, n (%)	
Not enrolled	188 (41.0)
Four-year college/university	186 (40.6)
Two-year community/junior college	55 (12.0)
High school	12 (2.6)
Technical school	10 (2.2)
Job training program	5 (1.1)
GED[a] classes	2 (0.4)
Relationship status, n (%)	
Single	215 (46.9)
Married	90 (19.6)
Divorced	4 (0.9)
In a serious relationship	149 (32.5)
Substance use, mean (SD)	
SPS[b] score (past year)	2.56 (4.23)
AUDIT[c] score	6.38 (5.69)

[a]GED: general equivalency diploma.

[b]SPS: Substance Problem Scale.

[c]AUDIT: Alcohol Use Disorder Identification Test.

Table 2. Peer Significant Other Behavior Questionnaire subscale responses to friend's substance use (n=208).

SBQ[a] item and subscale (1=highly unlikely, 5=highly likely)	Mean (SD)
Do you…	
stop your friend's drinking and drug use by getting angry (PSU[b])	2.07 (0.78)
have arguments regarding drinking and drug use when your friend is using (PSU)	2.62 (0.85)
show how unhappy you are when your friend is drinking or using drugs (PSU)	3.49 (0.82)
scare your friend from drinking or using drugs by warning (PSU)	2.29 (0.83)
tell your friend about things they did when drunk or high (PSU)	3.87 (0.78)
spend more time with your friend when they are not drinking or using (SS[c])	3.73 (0.83)
do things your friend likes when your friend is not drinking or using (SS)	3.96 (0.67)
give your friend compliments when they are not drinking or using (SS)	3.79 (0.77)
enjoy your friend's company when they are not drinking or using (SS)	4.21 (0.67)
do things for your friend when they are not drinking or using (SS)	3.93 (0.67)
support your friend when he or she is having trouble staying sober (SS)	4.06 (0.68)
arrange nondrinking or using social outings (SS)	3.79 (0.76)
when your friend is not drinking or using, do you enjoy each other (SS)	4.38 (0.52)
when your friend is drinking or using, do you join him/her (SSU[d])	2.62 (0.87)
bring alcoholic beverages or drugs to your friend's house (SSU)	1.94 (0.77)
buy alcohol for your friend when you are at bars, etc (SSU)	1.97 (0.77)
do things your friend likes when your friend is drinking or using (SSU)	2.67 (0.84)
tell your friend it is ok to have "just two or three" (SSU)	2.53 (0.89)
tell your friend they are fun to be with when they are drinking or using (SSU)	1.97 (0.71)
leave social situations when your fiend is drinking or using (WFF[e])	3.02 (0.83)
refuse to be around your friend when they are drinking or using (WFF)	2.91 (0.89)
keep out of your friend's way when he or she is drunk or high (WFF)	3.16 (0.82)
make other plans to go out alone or with others when your friend has been drinking or using (WFF)	3.51 (0.81)
refuse to take care of your friend when they are drunk or high (WFF)	2.26 (0.82)
Subscale	
Punish substance use	2.87 (0.55)
Support sobriety	3.98 (0.51)
Support substance use	2.28 (0.60)
Withdraw from friend	2.98 (0.67)

[a]SBQ: Significant Other Behavior Questionnaire.

[b]PSU: punish substance use.

[c]SS: support sobriety.

[d]SSU: support substance use.

[e]WFF: withdraw from friendship.

Hypothesis 1: Market Size/Acceptability

Approximately half of the participants (208/458, 45.4%) indicated that they had a friend experiencing problems with alcohol or drugs. A majority of the sample (334/458, 72.9%) reported they would be *somewhatlikely* or *completely likely* to use the mobile app, although the likelihood of use varied depending on the referral source. Fewer participants (222/458, 48.5%) reported that they would be at least *somewhat likely* to

search app stores on their own for such an app. However, if their close friend or someone directly associated with them (eg, their therapist, another close friend, a family member) asked them to use the app, 81.2% *(*372/458) of participants reported they would be at least *somewhat likely* to use it.

Hypothesis 2: Participant Intervention Script Ratings

Overall, participants rated the intervention scripts favorably in terms of linguistic/cultural accuracy and overall script utility.

XSL•FO
RenderX

Participants comprehensively rated the scripts as *slightly realistic* or better overall (mean 3.92, SD 0.48). Regarding utility, an overwhelming majority of peers (418/458, 91.3%) rated the scripts as at least slightly helpful in assisting them to support their friends who use substances. Participants identified scripts specifically designed to be enabling (mean 4.42, SD 0.49) and confrontational (mean 4.28, SD 0.50) as such. However, participants rated supportive scripts more neutrally (mean 2.96, SD 0.62).

Hypothesis 3: Participant Intervention Script Ratings and Self-Reported Interactions With Friends

Table 3 displays correlations between SBQ subscales (ie, self-reported actual behaviors) and confrontational, enabling, and supportive script ratings. Overall, the sample had higher scores for both enable drinking and confront drinking subscales relative to prior studies [43,44]. As hypothesized, there was a significant negative correlation ($r_{206}=-0.236$, $P=.001$) between enabling script ratings and scores on the SSU subscale of the SBQ, indicating that those who reported more enabling

behaviors in real life rated the intervention script dialogue as less enabling. Similarly, there was a significant positive correlation between enabling script ratings and scores on the SS subscale of the SBQ ($r_{206}=0.375$, $P<.001$).

As hypothesized, a positive correlation ($r=0.170$, $P=.01$) existed between supporting sobriety and confrontational script ratings, indicating that EAs who already support their friend's sobriety were more successful in identifying confrontational language. Conversely, correlations between supportive script ratings and the SS subscale were nonsignificant. Furthermore, nonsignificant correlations emerged between confrontational script ratings and the PSU subscale of the SBQ.

Participants rated to what degree they perceived these three scripts as supportive, enabling, or confrontational.

Table 4 displays intercorrelations between the SBQ subscales. Most were in the expected directions based on prior theory and research. For instance, participant scores on the SS subscale correlated negatively with scores on the SSU subscale ($r=-0.207$, $P=.003$).

Table 3. Script ratings and Significant Other Behavior Questionnaire subscale correlations.

Script type	PSU[a]	SS[b]	SSU[c]	WFF[d]
Supportive script				
Correlation	0.054	0.086	*0.157* [e]	0.031
P value (two-tailed)	.44	.22	.02	.66
Participants, n	208	208	208	208
Enabling script				
Correlation	*0.154*	*0.375*	*−0.236*	*0.274*
P value (two-tailed)	.03	<.001	.001	<.001
Participants, n	208	458	458	458
Confrontational script				
Correlation	0.008	*0.170*	−0.071	0.060
P value (two-tailed)	.91	.01	.31	.39
Participants, n	208	208	208	208

[a]PSU: punish substance use.

[b]SS: support sobriety.

[c]SSU: support substance use.

[d]WFF: withdraw from friendship.

[e]Italicized data represent findings significant at a level of <.05.

Table 4. Significant Other Behavior Questionnaire subscale intercorrelations.

SBQ subscale	PSU[a]	SS[b]	SSU[c]	WFF[d]
PSU				
Correlation	1	*0.406* [e]	0.004	*0.287*
P value (two-tailed)	__[f]	<.001	.94	<.001
Participants, n	208	208	208	208
SS				
Correlation	*0.406*	1	*−0.207*	*0.452*
P value (two-tailed)	<.001	—	<.001	<.001
Participants, n	208	208	208	208
SSU				
Correlation	0.004	*−0.207*	1	*−0.389*
P value (two-tailed)	.94	.003	—	<.001
Participants, n	208	208	208	208
WWF				
Correlation	*0.287*	*0.452*	*−0.389*	1
P value (two-tailed)	<.001	<.001	<.001	—
Participants, n	208	208	208	208

[a]PSU: punish substance use.

[b]SS: support sobriety.

[c]SSU: support substance use.

[d]WFF: withdraw from friendship.

[e]*Italicized* data represent findings significant at a level of <.05.

[f]Data not applicable (1:1 correlations).

Individuals who scored higher on the SS subscale of the SBQ also rated the scripts as more accurate overall in terms of the language used being representative of how EAs may text or converse with one another in more natural contexts (*r*=0.136, *P*=.001). Conversely, EAs who scored higher on the SS subscale of the SBQ rated the scripts as less useful overall (Table 5; *r*=−0.144, *P*=.04). Table 5 presents the correlations for the SBQ subscales with mean script accuracy and utility ratings across all scripts. Finally, aligned with a priori hypotheses, enabling script ratings were negatively correlated with AUDIT scores (*r*=−0.139, *P*=.003). In other words, participants with higher AUDIT scores rated enabling scripts lower (ie, lower perceived enabling).

Table 6 displays the correlations between the AUDIT scores and intervention script ratings.

This suggests that those with higher AUDIT scores may perceive their own substance use and attitudes toward their friends' substance use as nonenabling and similarly be an example of a subgroup that may benefit from using the Harbor app.

Table 5. Significant Other Behavior Questionnaire subscale and overall script accuracy and utility ratings.

SBQ[a] subscale	Script accuracy	Script utility
PSU[b]		
Pearson correlation	0.073	0.010
P value (two-tailed)	.29	.88
Participants, n	208	208
SS[c]		
Pearson correlation	*0.204* [d]	*−0.144*
P value (two-tailed)	.003	.04
Participants, n	208	208
SSU[e]		
Pearson correlation	−0.067	−0.010
P value (two-tailed)	.34	.89
Participants, n	208	208
WFF[f]		
Pearson correlation	0.021	0.035
P value (two-tailed)	.77	.61
Participants, n	208	208

[a]SBQ: Significant Other Behavior Questionnaire.

[b]PSU: punish substance use.

[c]SS: support sobriety.

[d]Italicized data represent findings significant at a level of <.05.

[e]SSU: support substance use.

[f]WFF: withdraw from friendship.

Table 6. Alcohol Use Disorder Identification Test score and script rating correlations.

Substance use measure	Confrontational script	Enabling script	Supportive script
AUDIT[a] score			
Pearson correlation	−0.020	*−0.139*[b]	*0.107*
P value (two-tailed)	.66	.003	.02
Participants, n	458	458	458

[a]AUDIT: Alcohol Use Disorder Identification Test.

[b]*Italicized* data represent findings significant at a level of <.05.

Discussion

Principal Findings

This study investigated 3 research questions: What is the potential market size of EAs who may use an app such as Harbor? What do potential Harbor users think of our intervention scripts? How do script ratings associate with self-reported participant behaviors toward existing friends' substance use problems? Systematic reviews on the functionality, feasibility, and acceptability of smartphone apps for EAs reveal a need for more rigorous evaluations of the perceived utility of and demand for apps [45]. From 2015 to 2019, smartphone ownership among EAs increased from 85% to 96% [46,47]. Considering this,

reaching EAs with substance use problems through technology is an attractive option.

Almost half of our sample (208/458, 45.4%) reported having a friend for whom they currently had concerns about their substance use, and 81.4% (373/458) of our survey respondents indicated they would be likely to download and use the Harbor app depending upon the referral source. This willingness to use such a program suggests a strong demand for a mobile app such as Harbor, as many EAs may have friends for whom they have concerns regarding substance use but lack the training and/or confidence to broach such issues with them. In addition, many participants in our sample were able to identify the supportive, confrontational, or enabling language within the scripts. Overall,

these findings indicate that the app is acceptable to participants, and interpretations may vary depending on the user.

Although participants' script utility ratings indicated some ambivalence about how helpful the app's scripts would be, the pattern of correlations makes a theoretical sense, that is, many of the correlations between the SBQ and intervention script ratings indicate that participants would have room for growth in communications with their EA peers affected with substance use. For instance, enabling script ratings correlated negatively with self-reported enabling behaviors in real life. This suggests that EAs' peers who self-reported enabling behaviors rate enabling scripts lower. This may indicate that decreasing enabling is a viable change target among such peers. Such findings may also support the construct validity of the single item used to measure the enabling script.

Conversely, those who already understand the concept of the role of enabling behavior in contributing to their peers' substance use may not benefit from the training module. The positive correlations between the SS subscale and both the enabling and confrontation script ratings underscore this point. Those who already use science-based communication and support their peers' sobriety likely recognize enabling and confrontational behaviors when they see them. For such individuals, the app may not teach new communication skills but rather affirm the skills they already have. This may help explain the negative correlation between supporting sobriety scores and script utility ratings (r_{206}=−0.097, P=.04). Although they rated the scripts as linguistically accurate, individuals who are already engaging in supportive behaviors seem to perceive their lack of personal need for such an app. The negative correlation could also indicate that those low on supportive behaviors rated the communication skills module as more useful, affirming its need among those for whom the content appeared novel.

Furthermore, these associations generally support the validity of these constructs among EAs. There are few studies on close friend communication patterns about substance use in the context of one peer having elevated use. To our knowledge, this is only the second study of SBQ measures involving EAs, and both support the notion that enabling, confrontation, and supporting sobriety are relevant concepts for peers with friends who misuse substances. It is encouraging that these associations support the theory-based (ie, CRAFT) intervention. Future research on Harbor could test whether confrontation, enabling, and support change over time for individuals who use the app.

Finally, it may be that the lower-than-expected ratings on the utility of the app were due to only showing select content of the app, that is, participants in this study did not see the entire intervention. Consequently, although they received a brief orientation to the purpose of the app before the presentation of the mock scripts, a lack of context about the rationale for speaking with friends about substance use in a specific way may have affected ratings. This study demonstrates some initial support for the acceptability and need for the intervention. However, future research on satisfaction with the Harbor app should come from actual users of the mobile app exposed to all content.

Limitations

Researchers and practitioners should examine the insights gained from this study in conjunction with the study's limitations. First, although the sample size is larger than or comparable to other feasibility studies, including technology-assisted interventions [48-51], it still may not be generalizable to the general EA population. It is also unclear whether peer reports of willingness to use the mobile app would translate to an increase in their supportive behaviors, and ultimately, a reduction in their close friend's substance use. Another potential limitation is the use of MTurk and self-reported data; however, according to prior research [52,53], both online survey data collection and self-reported data have demonstrated validity and reliability as research methods. To strengthen the reliability of the data gained from this study, researchers embedded several validity check items within the online survey. Furthermore, it should be noted that participants in this study did not download and interact with the app; they merely saw static screenshots and rated some of the scripts from 1 of 4 modules. Furthermore, additional engagement with the CSOs throughout intervention testing could build upon the CBPR framework used to guide this initial study. At the time of writing this paper, we have completed qualitative interviews with participants in the face-to-face version of the intervention. We did not verify survey findings with respondents, which would have enhanced participant involvement.

Future Studies

Mobile apps can make behavioral health care more accessible, interactive, and efficient [15]. In general, clients find digital interventions easy to operate and useful, and rates of prolonged use are generally high [54]. This study demonstrated the acceptability and feasibility of Harbor, a mobile app designed to teach EAs who have friends with substance use problems ways to intervene with their friends' substance use. Substance use in EAs is a major public health problem [55], and Harbor may be useful in extending both the reach and efficacy of treatments targeting this population. The findings also demonstrated the construct validity of enabling, confronting, and supporting behaviors among peers of EAs who may be struggling with substance use problems. Future research on Harbor can test its efficacy in improving EAs' substance use treatment outcomes by altering these peer behaviors.

XSL·FO

RenderX

Conflicts of Interest

DS is the lead developer of the Harbor app, which is owned by the University of Illinois' Board of Trustees.

Multimedia Appendix 1
Screenshot of example confrontational script and associated questionnaire items.
[PNG File , 963 KB - jmir_v22i11e16632_app1.png]

Multimedia Appendix 2
Screenshot of example enabling script and associated response item.
[PNG File , 880 KB - jmir_v22i11e16632_app2.png]

Multimedia Appendix 3
Screenshot of supportive script and associated response item.
[PNG File , 1664 KB - jmir_v22i11e16632_app3.png]

References

1. Behavioral Health Trends in the United States: Results From the 2014 National Survey on Drug Use and Health. The Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf [accessed 2020-09-17]

2. Adams SH, Knopf DK, Park MJ. Prevalence and treatment of mental health and substance use problems in the early emerging adult years in the United States. Emerg Adulthood 2013 Dec 2;2(3):163-172. [doi: 10.1177/2167696813513563]

3. Reach of Selected Social Networks in the United States as of February 2017, by Age Group. Statista. 2017. URL: https://www.statista.com/statistics/305245/us-social-network-penetration-age-group/ [accessed 2020-09-17]

4. Behrend TS, Sharek DJ, Meade AW, Wiebe EN. The viability of crowdsourcing for survey research. Behav Res Methods 2011 Sep;43(3):800-813. [doi: 10.3758/s13428-011-0081-0] [Medline: 21437749]

5. Susukida R, Crum RM, Ebnesajjad C, Stuart EA, Mojtabai R. Generalizability of findings from randomized controlled trials: application to the national institute of drug abuse clinical trials network. Addiction 2017 Jul;112(7):1210-1219 [FREE Full text] [doi: 10.1111/add.13789] [Medline: 28191694]

6. Mathews A, Farley S, Hightow-Weidman L, Muessig K, Rennie S, Tucker JD. Crowdsourcing and community engagement: a qualitative analysis of the 2BeatHIV contest. J Virus Erad 2018 Jan 1;4(1):30-36 [FREE Full text] [Medline: 29568551]

7. Meyers R, Smith J. Clinical Guide to Alcohol Treatment: The Community Reinforcement Approach. New York, USA: Guilford Press; 1995.

8. Prior L. Belief, knowledge and expertise: the emergence of the lay expert in medical sociology. Sociol Health Illn 2003;25:41-57. [doi: 10.1111/1467-9566.00339] [Medline: 14498929]

9. Crawford R. Health as a meaningful social practice. Health (London) 2006 Oct;10(4):401-420. [doi: 10.1177/1363459306067310] [Medline: 16973678]

10. Lane K. The plasticity of professional boundaries: a case study of collaborative care in maternity services. Health Soc Rev 2014 Dec 17;15(4):341-352. [doi: 10.5172/hesr.2006.15.4.341]

11. Tousijn W. Beyond decline: consumerism, managerialism and the need for a new medical professionalism. Health Soc Rev 2014 Dec 17;15(5):469-480. [doi: 10.5172/hesr.2006.15.5.469]

12. Stewart M, Brown J, Weston W, McWhinney I, McWilliam C, Freeman T. Patient-Centered Medicine: Transforming the Clinical Method. New York, USA: CRC Press; 2013.

13. Elwyn G, Edwards A, Kinnersley P, Grol R. Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. Br J Gen Pract 2000 Nov;50(460):892-899. [Medline: 11141876]

14. Mazur D. The New Medical Conversation: Media, Patients, Doctors, and the Ethics of Scientific Communication. New York, USA: Rowman & Littlefield; 2003.

15. Henderson J. Expert and lay knowledge: A sociological perspective. 2010;67(1):A. [doi: 10.1111/j.1747-0080.2010.01409.x]

16. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]

17. Bertholet N, Cunningham JA, Faouzi M, Gaume J, Gmel G, Burnand B, et al. Internet-based brief intervention to prevent unhealthy alcohol use among young men: a randomized controlled trial. PLoS One 2015;10(12):e0144146 [FREE Full text] [doi: 10.1371/journal.pone.0144146] [Medline: 26642329]

18. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health 2010 Apr 1;100(Suppl 1):S40-S46. [doi: 10.2105/AJPH.2009.184036] [Medline: 20147663]

19. Spijkerman R, Roek MA, Vermulst A, Lemmers L, Huiberts A, Engels RC. Effectiveness of a web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. J Med Internet Res 2010 Dec 19;12(5):e65 [FREE Full text] [doi: 10.2196/jmir.1465] [Medline: 21169172]

20. Hacker K, Tendulkar SA, Rideout C, Bhuiya N, Trinh-Shevrin C, Savage CP, et al. Community capacity building and sustainability: outcomes of community-based participatory research. Prog Community Health Partnersh 2012;6(3):349-360 [FREE Full text] [doi: 10.1353/cpr.2012.0048] [Medline: 22982848]

21. Smith DC, Cleeland L, Middleton A, Godley MD. Willingness and appropriateness of peers participating in emerging adults' substance misuse treatment. J Subst Abuse Treat 2013 Jul;45(1):148-154 [FREE Full text] [doi: 10.1016/j.jsat.2013.01.008] [Medline: 23462152]

22. Smith DC, Davis JP, Ureche DJ, Dumas TM. Six month outcomes of a peer-enhanced community reinforcement approach for emerging adults with substance misuse: a preliminary study. J Subst Abuse Treat 2016 Feb;61:66-73 [FREE Full text] [doi: 10.1016/j.jsat.2015.09.002] [Medline: 26482135]

23. Lewis MA, Neighbors C. Social norms approaches using descriptive drinking norms education: a review of the research on personalized normative feedback. J Am Coll Health 2006;54(4):213-218 [FREE Full text] [doi: 10.3200/JACH.54.4.213-218] [Medline: 16450845]

24. Davis JP, Smith DC, Briley DA. Substance use prevention and treatment outcomes for emerging adults in non-college settings: a meta-analysis. Psychol Addict Behav 2017 May;31(3):242-254. [doi: 10.1037/adb0000267] [Medline: 28318279]

25. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011 Dec;42(6):505-512. [doi: 10.1037/a0024485]

26. Rouse SV. A reliability analysis of Mechanical Turk data. Comput Hum Behav 2015 Feb;43:304-307. [doi: 10.1016/j.chb.2014.11.004]

27. Casler K, Bickel L, Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Comput Hum Behav 2013 Nov;29(6):2156-2160. [doi: 10.1016/j.chb.2013.05.009]

28. Zhang C, Conrad F. Speeding in web surveys: the tendency to answer very fast and its association with straightlining. Res Methods 2014;8(2):127-135 [FREE Full text]

29. Love CT, Longabaugh R, Clifford PR, Beattie M, Peaslee CF. The Significant-other behavior questionnaire (SBQ): an instrument for measuring the behavior of significant others towards a person's drinking and abstinence. Addiction 1993 Sep;88(9):1267-1279. [doi: 10.1111/j.1360-0443.1993.tb02149.x] [Medline: 8241926]

30. Rychtarik RG, McGillicuddy NB. The spouse situation inventory: a role-play measure of coping skills in women with alcoholic partners. J Fam Psychol 1997 Sep;11(3):289-300 [FREE Full text] [doi: 10.1037/0893-3200.11.3.289] [Medline: 16981014]

31. Walitzer KS, Dermen KH. Alcohol-focused spouse involvement and behavioral couples therapy: evaluation of enhancements to drinking reduction treatment for male problem drinkers. J Consult Clin Psychol 2004 Dec;72(6):944-955. [doi: 10.1037/0022-006X.72.6.944] [Medline: 15612842]

32. Kelly T, Donovan J, Kinnane J, Taylor D. A comparison of alcohol screening instruments among under-aged drinkers treated in emergency departments. Alcohol Alcohol 2002;37(5):444-450. [doi: 10.1093/alcalc/37.5.444] [Medline: 12217936]

33. Kelly T, Donovan J, Chung T, Cook R, Delbridge T. Alcohol use disorders among emergency department-treated older adolescents: a new brief screen (RUFT-Cut) using the AUDIT, CAGE, CRAFFT, and RAPS-QF. Alcohol Clin Exp Res 2004 May;28(5):746-753. [doi: 10.1097/01.alc.0000125346.37075.85] [Medline: 15166649]

34. Kokotailo P, Egan J, Gangnon R, Brown D, Mundt M, Fleming M. Validity of the alcohol use disorders identification test in college students. Alcohol Clin Exp Res 2004 Jun;28(6):914-920. [doi: 10.1097/01.alc.0000128239.87611.f5] [Medline: 15201634]

35. American PA. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Washington, DC: Author; 2013.

36. Smith DC, Cleeland L, Dennis ML. Reasons for quitting among emerging adults and adolescents in substance-use-disorder treatment. J Stud Alcohol Drugs 2010 May;71(3):400-409 [FREE Full text] [doi: 10.15288/jsad.2010.71.400] [Medline: 20409434]

37. Smith DC, Godley SH, Godley MD, Dennis ML. Adolescent community reinforcement approach outcomes differ among emerging adults and adolescents. J Subst Abuse Treat 2011 Dec;41(4):422-430 [FREE Full text] [doi: 10.1016/j.jsat.2011.06.003] [Medline: 21831564]

38. Smith DC, Bahar OS, Cleeland LR, Davis JP. Self-perceived emerging adult status and substance use. Psychol Addict Behav 2014 Sep;28(3):935-941 [FREE Full text] [doi: 10.1037/a0035900] [Medline: 25134032]

39. Smith DC, Bennett KM, Dennis ML, Funk RR. Sensitivity and specificity of the gain short-screener for predicting substance use disorders in a large national sample of emerging adults. Addict Behav 2017 May;68:14-17. [doi: 10.1016/j.addbeh.2017.01.013] [Medline: 28088053]

40. Chan Y, Dennis ML, Funk RR. Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment. J Subst Abuse Treat 2008 Jan;34(1):14-24 [FREE Full text] [doi: 10.1016/j.jsat.2006.12.031] [Medline: 17574804]

XSL•FO
RenderX

41. IBM SPSS Statistics for Windows. IBM Corp. 2017. URL: https://www.ibm.com/in-en/analytics/spss-statistics-software [accessed 2020-09-24]

42. The MacArthur Scale of Subjective Social Status. 2007. URL: http://sparqtools.org/mobility-measure/macarthur-scale-of-subjective-social-status-adult-version/ [accessed 2020-09-17]

43. Rodriguez LM, Osilla KC, Trail TE, Gore KL, Pedersen ER. Alcohol use among concerned partners of heavy drinking service members and veterans. J Marital Fam Ther 2018 Apr;44(2):277-291 [FREE Full text] [doi: 10.1111/jmft.12261] [Medline: 28782116]

44. Monti PM, Colby SM, Mastroleo NR, Barnett NP, Gwaltney CJ, Apodaca TR, et al. Individual versus significant-other-enhanced brief motivational intervention for alcohol in emergency care. J Consult Clin Psychol 2014 Dec;82(6):936-948 [FREE Full text] [doi: 10.1037/a0037658] [Medline: 25111430]

45. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]

46. Smith A. A Portrait of Smartphone Ownership. Pew Research Center. 2015. URL: http://www.pewinternet.org/2015/04/01/chapter-one-a-portrait-of-smartphone-ownership/ [accessed 2020-09-17]

47. Mobile fact sheet. Pew Research Center. 2019. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-09-17]

48. Bersamin M, Paschall MJ, Fearnow-Kenney M, Wyrick D. Effectiveness of a web-based alcohol-misuse and harm-prevention course among high- and low-risk students. J Am Coll Health 2007;55(4):247-254. [doi: 10.3200/JACH.55.4.247-254] [Medline: 17319331]

49. Bewick BM, Trusler K, Mulhern B, Barkham M, Hill AJ. The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: a randomised control trial. Addict Behav 2008 Sep;33(9):1192-1198. [doi: 10.1016/j.addbeh.2008.05.002] [Medline: 18554819]

50. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a Facebook smoking cessation intervention for young adults. J Med Internet Res 2015 Dec 31;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]

51. Ruggiero KJ, Resnick HS, Acierno R, Coffey SF, Carpenter MJ, Ruscio AM, et al. Internet-based intervention for mental health and substance use problems in disaster-affected populations: a pilot feasibility study. Behav Ther 2006 Jun;37(2):190-205. [doi: 10.1016/j.beth.2005.12.001] [Medline: 16942971]

52. Chan D. So why ask me? Are self-report data really that bad? In: Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences. New York, USA: Routledge/Taylor & Francis Group; 2009.

53. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]

54. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]

55. Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2016. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR1-2016/NSDUH-FFR1-2016.htm [accessed 2020-09-24]

Abbreviations

AUDIT: Alcohol Use Disorder Identification Test
CBPR: community-based participatory research
CRAFT: Community Reinforcement Approach With Family Training
CSOs: concerned significant others
EA: emerging adult
MTurk: Amazon's Mechanical Turk
Peer-CRA: peer-enhanced community reinforcement approach
PSU: punish substance use
SBQ: Significant Other Behavior Questionnaire
SPS: Substance Problem Scale
SS: support sobriety
SSU: support substance use

XSL•FO
RenderX

Original Paper

Identifying Persuasive Design Principles and Behavior Change Techniques Supporting End User Values and Needs in eHealth Interventions for Long-Term Weight Loss Maintenance: Qualitative Study

Rikke Aune Asbjørnsen[1,2,3], MSc; Jobke Wentzel[1,4], PhD; Mirjam Lien Smedsrød[5], MSc; Jøran Hjelmesæth[6,7], PhD; Matthew M Clark[8], PhD; Lise Solberg Nes[3,8,9*], PhD; Julia E W C Van Gemert-Pijnen[1,10,11*], PhD

[1]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[2]Research and Innovation Department, Vestfold Hospital Trust, Tønsberg, Norway

[3]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway

[4]Research Group IT Innovations in Health Care, Windesheim University of Applied Sciences, Zwolle, Netherlands

[5]Collaborative Care Unit, Sørlandet Hospital Trust, Kristiansand, Norway

[6]Morbid Obesity Center, Vestfold Hospital Trust, Tønsberg, Norway

[7]Department of Endocrinology, Morbid Obesity and Preventive Medicine, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[8]Department of Psychiatry & Psychology, Mayo Clinic, College of Medicine & Science, Rochester, MN, United States

[9]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[10]University Medical Center Groningen, Groningen, Netherlands

[11]University of Waterloo, Waterloo, ON, Canada

*these authors contributed equally

Corresponding Author:
Rikke Aune Asbjørnsen, MSc
Centre for eHealth and Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
De Zul 10
Enschede, 7522 NJ
Netherlands
Phone: 31 534899111
Email: r.a.asbjornsen@utwente.nl

Abstract

Background: An increasing number of eHealth interventions aim to support healthy behaviors that facilitate weight loss. However, there is limited evidence of the effectiveness of the interventions and little focus on weight loss maintenance. Knowledge about end user values and needs is essential to create meaningful and effective eHealth interventions, and to identify persuasive system design (PSD) principles and behavior change techniques (BCTs) that may contribute to the behavior change required for successful long-term weight loss maintenance.

Objective: This study aimed to provide insight into the design of eHealth interventions supporting behavior change for long-term weight maintenance. The study sought to identify the values and needs of people with obesity aiming to maintain weight after weight loss, and to identify PSD principles, BCTs, and design requirements that potentially enable an eHealth intervention to meet end user values and needs.

Methods: This study presents the concept of integrating PSD principles and BCTs into the design process of eHealth interventions to meet user values and needs. In this study, individual interviews and focus groups were conducted with people with obesity (n=23) and other key stakeholders (n=27) to explore end user values and needs related to weight loss maintenance. Design thinking methods were applied during the focus group sessions to identify design elements and to explore how eHealth solutions can support the needs to achieve sustainable weight loss maintenance. The PSD model and behavior change taxonomy by Michie were used to identify PSD principles and BCT clusters to meet end user values and needs.

XSL•FO
RenderX

Results: A total of 8 key end user values were identified, reflecting user needs for weight loss maintenance support: *self-management*, *personalized care*, *autonomy*, *feel supported*, *positive self-image*, *motivation*, *happiness*, and *health*. *Goals and planning*, *feedback and monitoring*, *repetition and substitution*, *shaping knowledge*, *social support*, *identity*, and *self-belief* were some of the BCT clusters identified to address these concepts, together with PSD principles such as *personalization*, *tailoring*, *self-monitoring*, *praise*, and *suggestions*.

Conclusions: The process of translating end user values and needs into design elements or features of eHealth technologies is an important part of the design process. To our knowledge, this is the first study to explore how PSD principles and BCTs can be integrated when designing eHealth self-management interventions for long-term weight loss maintenance. End users and other key stakeholders highlighted important factors to be considered in the design of eHealth interventions supporting sustained behavior change. The PSD principles and BCTs identified provide insights and suggestions about design elements and features to include for supporting weight loss maintenance. The findings indicate that a combination of BCTs and PSD principles may be needed in evidence-based eHealth interventions to stimulate motivation and adherence to support healthy behaviors and sustained weight loss maintenance.

Trial Registration: ClinicalTrials.gov NCT04537988; https://clinicaltrials.gov/ct2/show/NCT04537988

Introduction

The Challenges of Weight Loss Maintenance

Globally, obesity has grown to epidemic proportions and nearly tripled between 1975 and 2016 [1]. In 2016, more than 650 million adults worldwide were obese [1], which led to individual and societal challenges with regard to health, well-being, and economic burden [2-5]. Many people with obesity manage to lose weight, but an alarmingly large percentage of people (3 out of 4) fail to maintain the lost weight over time [6]; therefore, long-term weight maintenance following weight loss is a major concern.

Sustained health behavior change is required to prevent weight regain [7-9], and the chance of success increases when individuals maintain healthy behaviors and weight loss for 2 to 5 years [10,11]. The US National Weight Control Registry [12] and several weight maintenance–related studies provide information about factors [13] associated with successful long-term weight maintenance [8-12]. The identified factors include continued adherence to behavioral strategies such as frequent self-weighing and habitual routines, including high levels of physical activity, healthy diet, eating breakfast regularly, and consistent eating patterns across weekdays and weekends [8,10,13-20].

The challenges that people face when trying to prevent weight regain after weight loss are complex, with several biological, behavioral, cognitive, emotional, social, and environmental factors interacting [16,21]. To address these challenges, novel solutions and emerging technologies, such as eHealth solutions, have the potential to support the self-management and behavior change processes needed for continued weight control [17,21-26]. At present, several eHealth interventions are available to support weight loss [23,27-29]. However, evidence of long-term effects related to weight loss maintenance solutions is limited. Research examining the potential of design for sustained behavior change through eHealth interventions [27] focusing on weight loss maintenance is therefore needed [30,31].

Combining Persuasive and Behavior Change Techniques in eHealth Design

eHealth interventions and persuasive technologies offer possibilities for improving self-management of health and are increasingly used to support healthy behaviors for improved health and well-being [26,28-32]. Establishing habits for long-term behavior change and weight loss maintenance is challenging and takes time [10,33]. For long-term behavior change, eHealth interventions need to provide effective self-management strategies, support lifestyle change, and promote healthy behaviors. A major challenge to effective eHealth is to integrate motivating and engaging design elements for adherence and continuity of use [23,34,35]. Despite these challenges, such technologies have the potential to empower individuals and transform health care by shifting the focus from cure to prevention and self-management with improved health outcomes [36-38]. This requires not only attention to existing evidence from research and clinical practice but also careful design in collaboration with end users and other stakeholders to develop and implement feasible and sustainable eHealth solutions [36,39-42]. In the design of eHealth interventions, behavior change techniques (BCTs) [43] and persuasive system design (PSD) principles [44] are increasingly applied to design effective behavioral interventions that motivate, engage, and promote healthy lifestyles in support of continued behavior change [45-47]. Behavior change taxonomy by Michie [43] has been systematically developed to meet the need for standardized reporting when designing and evaluating complex behavior change interventions, building on a broad range of BCTs to describe intervention components and content, independent of theory. The PSD model by Oinas-Kukkonen [26] includes a range of PSD principles that can be applied when designing novel and user-friendly eHealth interventions and persuasive technology to support healthy behaviors and behavior change.

However, the most effective combination of such techniques and principles remains unclear [23,26,36,43,47-55].

The research team of this study had recently conducted a scoping review [23] that identified PSD principles [26] (eg, *personalization*, *self-monitoring*, *tailoring*, *praise*, *suggestions*, *rewards*, and *reminders*) and BCT clusters [43] (eg, *feedback and monitoring*, *goals and planning*, *social support*, *shaping knowledge*, *associations*, and *repetition* and *substitution*) applied in eHealth interventions to stimulate motivation, adherence, and weight loss maintenance [23]. Findings from the review suggest different strategies for losing weight than for weight maintenance [23]. Although eHealth interventions, when developed in line with user needs [36,56,57], have the potential to support the difficult behavior change process needed to prevent long-term weight regain [10,21,23,58-60], the scoping review revealed that user involvement in the design and development of such eHealth interventions is lacking [23]. Moreover, there is insufficient knowledge about the most ideal combinations of PSD principles and BCTs to support weight maintenance over time [23]. To create meaningful and effective digital behavior change interventions, it is essential to identify the end user values and needs and the related PSD principles and BCTs during the requirement specification process [23,32,36,39,61]. A holistic and broad stakeholder-driven approach, including key stakeholders, is imperative to design and develop sustainable eHealth interventions that reflect the values and support the goals of the end users [36,39].

Objectives

The overall aim of this study was to provide insight into the design of eHealth interventions aiming to support behavior change for long-term weight loss maintenance. The goal was to identify values and needs of people with obesity aiming to maintain weight after weight loss (ie, end users) and how these values and needs can be met by PSD principles and BCTs in eHealth intervention design. The following research questions were addressed:

1. What are the values and needs of people with obesity (ie, end users) aiming to maintain weight after initial weight loss, according to key stakeholders (eg, end users and health care providers)?
2. Which PSD principles, BCTs, and design requirements can be identified to meet end user values and needs?

End user *values* refer to the main drivers of behavior or high-level needs or requirements (eg, improved health and motivation to maintain healthy routines in the long term), indicating the added values of the technology or a solution to support users in reaching their goals (ie, maintaining weight after weight loss) [36,39]. End user *needs* refer to demands and low-level requirements (ie, low-level needs) that people want to solve to address their problems (eg, realistic goal setting, self-regulation, and coping skills to maintain weight) [36,39]. A value can drive several underlying needs [36,39,62].

A *key stakeholder* was defined as someone central in understanding end user challenges and needs, including possible latent and future needs, and someone that could provide valuable input on how to meet end user needs in eHealth technology design [36,63].

Methods

The Double Diamond Framework and the Center for eHealth Research and Disease Management Roadmap

This study builds on the innovation and design framework of the Double Diamond [64,65] and the Center for eHealth Research and Disease Management (CeHRes) comprehensive roadmap [36,39] to guide the design and development of an evidence-based eHealth intervention supporting long-term behavior change and weight loss maintenance. Figure 1 presents a combination of the Double Diamond [64] and the CeHRes roadmap [36,39].

Figure 1. The Double Diamond [64] and the Center for eHealth Research and Disease Management roadmap [39] combined: a design thinking process for eHealth design and development.
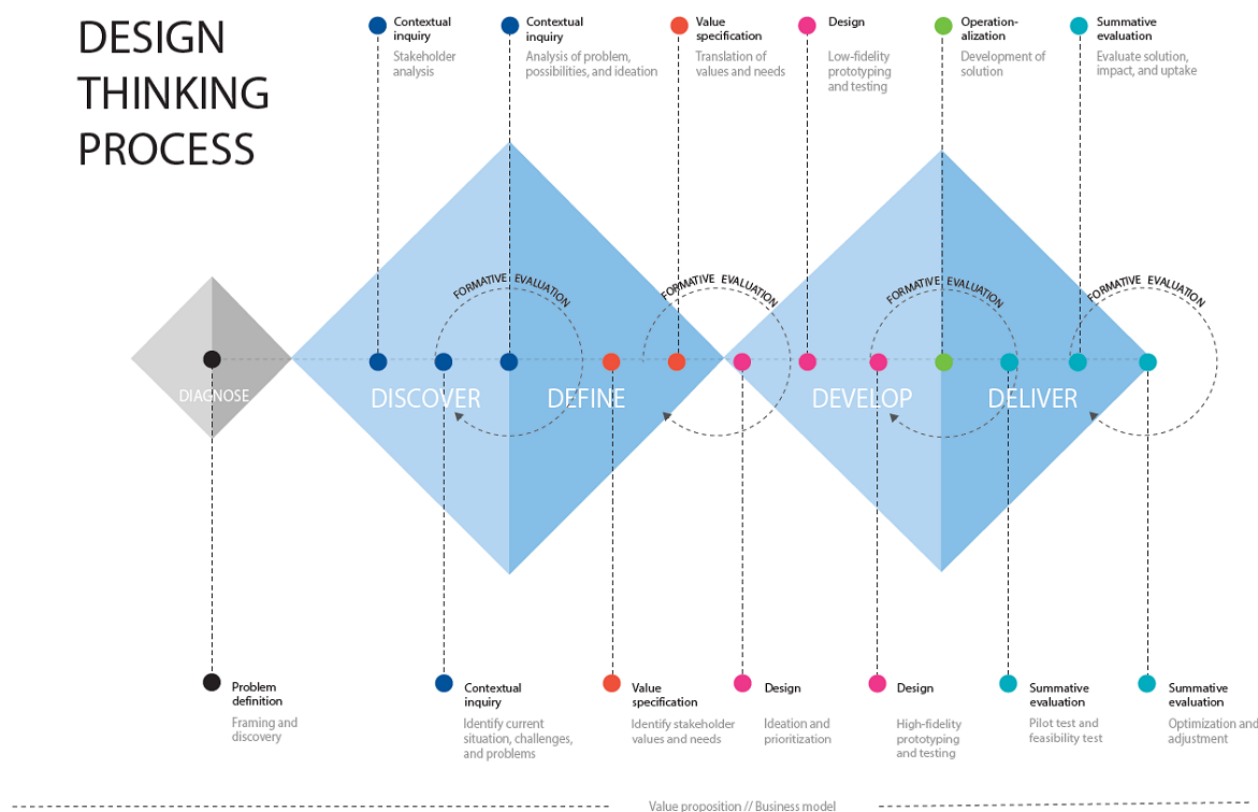
As a first step, a *diagnose phase* was included [66,67], consisting of the previously performed scoping review [23] and an alignment workshop with the research group (including a user representative) to define the issues at hand (ie, problem definition) and create an initial overview of the stakeholders in the weight loss maintenance conundrum.

The shape of the Double Diamond represents the process of gathering insights (diverging to discover) and translating these insights into ideas or concepts (converging to define) [21,64,65,68-72]. The Double Diamond framework utilizes design thinking methods that can be applied to identify user needs and to find innovative solutions for complex health care challenges [63,70,71,73]. Design thinking explores and elaborates on what is desirable from a user perspective, what is technologically feasible, and what is economically viable [74]. Complementing the design thinking process, the CeHRes roadmap [36,39] was applied as a guideline for holistic development, implementation, and evaluation of eHealth technologies. The CeHRes roadmap consists of the following iterative phases: *contextual inquiry*, *value specification*, *design*, *operationalization*, and *summative evaluation* [36,39]. The roadmap integrates concepts of persuasive technology design [26], human-centered design, and business modeling through participatory development with continuous evaluation cycles during the eHealth development and implementation process [39].

This study focuses on the *discover* and *define* phase [64] (Figure 1), including the *contextual inquiry* and *value specification* phases [39]. During the *discover* phase [64], a stakeholder analysis [63] was executed to identify key stakeholders who could provide insights into the needs, challenges and problems of people aiming to maintain weight after weight loss (ie, end users) [63]. As part of the *contextual inquiry*, individual interviews and focus group sessions were performed to elicit end user values and needs and to inquire how needs can be met by eHealth technology [36]. During the *define* phase [64], the values and needs of end users were identified, resulting in *value specification* [39]. Finally, the PSD model [26], behavior change taxonomy by Michie [43], and the results from the recent scoping review [23] were used to translate values and needs into PSD principles and BCTs [39,75,76].

Key Stakeholder Identification and Recruitment

Stakeholder Analysis

To identify key stakeholders, as part of the *discover* phase, a service design workshop [63] was organized with a multidisciplinary team of stakeholders, consisting of end user representatives (ie, people with obesity aiming to maintain weight after initial weight loss; n=2), health care personnel (ie, health care providers; n=5), and project group representatives (n=3). See Figure 2 for study participants, including the participants in the service design workshop.

Figure 2. Study participants and data collection.

Double Diamond framework	Research goal and questions (RQ)	Method	End users[a]	Other key stakeholders
DISCOVER	Identifying key stakeholders (to answer in RQ1)	Service design workshop	End user representative (n=2)	eHealth researcher (n=1) Clinical health psychologist (n=1) eHealth content developer (n=1) Obesity experts (i.e., physician, clinical nutritionist, nurse, and behavioral scientist) (n=5)
	Identifying values & needs of primary end-users (to answer RQ1)	Individual interviews	End users (n=10) Patient organization/ end user representative (n=1)	Primary health care providers[b] (n=7) Secondary health care providers (Obesity specialist centers) (n=4) Private sector (n=1) Policy level (n=1)
DEFINE	Identifying values, needs & design elements to support weight maintenance (to answer RQ1 and RQ2)	Focus group interviews	End users (n=10)[c]	Secondary health care providers (i.e., physician, behavioral therapist, clinical nutritionists, and exercise physiologist) (n=6)
		Total N=50	n=23	n=27

[a]End users: people with obesity aiming to maintain weight after initial weight loss.

[b]One interview was performed as a group interview.

[c]A total of 2 end users (ie, patients) participated in both individual interviews and focus group

The stakeholder identification workshop was led by a service designer and aimed to create a map of key stakeholders [63] to involve in the research and design process. On the basis of the methods from stakeholder theory and service design [36,63,73,77,78], idea generation was facilitated during the workshop by using a large whiteboard stakeholder map as a mind mapping tool [63], generating a preliminary list of potential stakeholders. The initial list of stakeholders was then narrowed to key stakeholders based on group discussions and mutual agreement during the workshop session. The broad and multidisciplinary stakeholder involvement in this study is part of a holistic and human-centered development [36,39,70,73].

Recruitment of End Users and Other Key Stakeholders

Prospective *end users* (ie, people with obesity aiming to maintain weight after initial weight loss—the intervention target group) were defined as adults aged 18 years or above, with BMI ≥30 kg/m^2 [79] before weight loss, who had recently lost 8% weight or more (eg, through a low-calorie diet or lifestyle change program), and who were in need of support to prevent weight regain. People who met these criteria and had previously experienced weight loss, which was followed by weight regain and/or weight maintenance, were selected for study participation. Eligible participants also had to have access to the internet and be fluent in the Norwegian language. Recruitment was conducted over several weeks by 3 collaborative leading obesity research and treatment centers in Norway through convenience

sampling, that is, people meeting the inclusion criteria were invited to participate in the study by their health care provider when undergoing treatment or during outpatient follow-up visits. Potential study participants (ie, prospective end users) received written and oral study information, and were invited to participate in individual interviews and focus groups at local hospitals and research centers (Figure 2). If interested, the potential participants were contacted by the research team, provided with an appointment time, and informed that they would receive a gift certificate (approximate value US $30) as compensation for time spent or of potential costs (eg, transport and parking). Representatives of *other key stakeholders* identified during the stakeholder analysis (eg, health care providers) were contacted based on convenience sampling and invited to participate in individual interviews and focus groups through the collaborating health care institutions. See Figure 2 for details.

Ethical Approval and Informed Consent

This study was approved by the Hospital Privacy and Security Protection Committee (institutional review board, approval number: 2017/12702) at Oslo University Hospital (OUH) in Norway, one of the largest medical centers in northern Europe. All participants involved in this study received study information and signed a written informed consent form before participation.

Identification of End User Values and Needs

Semistructured Interviews

As part of the data collection, individual semistructured interviews were performed with prospective end users (n=11) who had recently lost weight and needed support and guidance to prevent weight regain. Other key stakeholders (n=13) were also interviewed to capture a broad perspective and understanding of end user challenges, values, and needs to maintain weight after weight loss [36,73,80] (Figure 2). Interviews with end users and other key stakeholders were guided by the overarching research question: *What are the values and needs of people with obesity aiming to maintain weight after initial weight loss?*

Individual interviews (60-min to 90-min long) were performed and audio recorded by the first and third author (RA: 17/23, 73%; MS: 6/23, 26%). A semistructured interview guide was developed (Multimedia Appendix 1), providing an overview of the themes to explore. These themes included everyday life, behavior, thoughts, feelings, routines, challenges, strategies, values and needs to maintain weight, experiences and preferences related to health apps, activity and nutrition tracker, and weight management technology (eg, apps, health forums, web-based weight management programs, videos, blogs, and wearables).

Focus Groups

To further explore and validate the everyday needs of end users and elaborate on engaging and motivating design elements that could potentially support weight maintenance in eHealth design, 3 focus group sessions were held [80-82]. In 2 focus groups, 10 end users (n=10) participated, 5 in each group. In the third focus group, other key stakeholders (ie, health care personnel) participated (n=6) [81,82] (Figure 2). The focus groups lasted approximately 2.5 hours and were facilitated by a service designer and the first author (RA). A digital designer and a Scrum (ie, Agile project management methodology or framework) product owner also participated in the focus groups. The focus group sessions had an explorative and creative character, applying participatory and ideation methods from design thinking [73]. All 3 focus groups were given the following assignment and topics to discuss.

First, a "What is your favorite app?" assignment was performed, where participants described and/or suggested their own apps. The assignment was performed to stimulate creativity and to create trust and social connections within the group. During the assignment, positive and negative aspects of the apps were elaborated on employing brainstorming techniques. The following topics were then explored to provide input for the main research questions:

- What do you/end users need to maintain weight?

- Which design elements can help you/end users to maintain weight?

Sticky notes and a large whiteboard were used to discuss these topics (ie, open-ended questions) with the focus group participants, eliciting potential unmet user needs. This technique was used to create an image of the diversity of user needs [73]. A dynamic and informal discussion, where focus group participants could instantly react to each other's suggestions, provided the researchers with a broad impression of the topics [81].

To identify engaging and motivating design elements, feature cards were applied based on design methods and principles from open card sorting (although applied in a less traditional way than card sorting tasks for information architecture; Multimedia Appendix 2) [83-87]. The cards contained PSD principles and BCTs identified in the previously mentioned scoping review [23]. Additional design elements were included to explore other potentially engaging elements supporting behavior change and to facilitate the creative process related to users' requirements, preferences, and ideas to meet their needs [23,86-90]. These elements include metaphors (eg, when a target is reached, a flower grows in a garden), avatars (eg, a virtual or digital person based on oneself), and gamification elements (eg, points and trophies).

The participants could also choose to develop their own cards with motivational elements through sketching or writing, as some cards had no content (ie, were blank). Examples of how the design elements could be used in app design were presented during the introduction. The focus group participants could choose and sort the cards they identified as useful (ie, to motivate and support weight maintenance) and then share their results in an open group discussion. Participants sorted the selected cards into groups of *must have*, *nice to have*, and *not needed*.

Translation of Values and Needs Into PSD Principles, BCTs, and Design Requirements

The requirement elicitation methods (ie, individual interviews, focus groups) aimed to capture the user perception of what end users need and how the design of an eHealth technology should be [36]. A total of 3 researchers (RA, MS, and JW) participated in the analysis and validation process of values and needs before translating these into PSD principles, BCTs, and high-level requirements for the design and development of an eHealth weight loss maintenance intervention [91] (Multimedia Appendix 3) [76]. Figure 3 illustrates the steps used when translating end user values and needs into PSD principles, BCT clusters, and high-level design requirements. The presented methods provided input for the main research question: *Which PSD principles, BCTs, and design requirements can be identified to meet the end user´ values and needs?*

Figure 3. Value identification and translation of identified values and needs into persuasive system designs, behavior change techniques, and high-level requirements. BCT: behavior change technique; PSD: persuasive system design.

*Identification of PSD principles and BCTs based on the PSD-model and Michie's taxonomy

**Findings from literature review
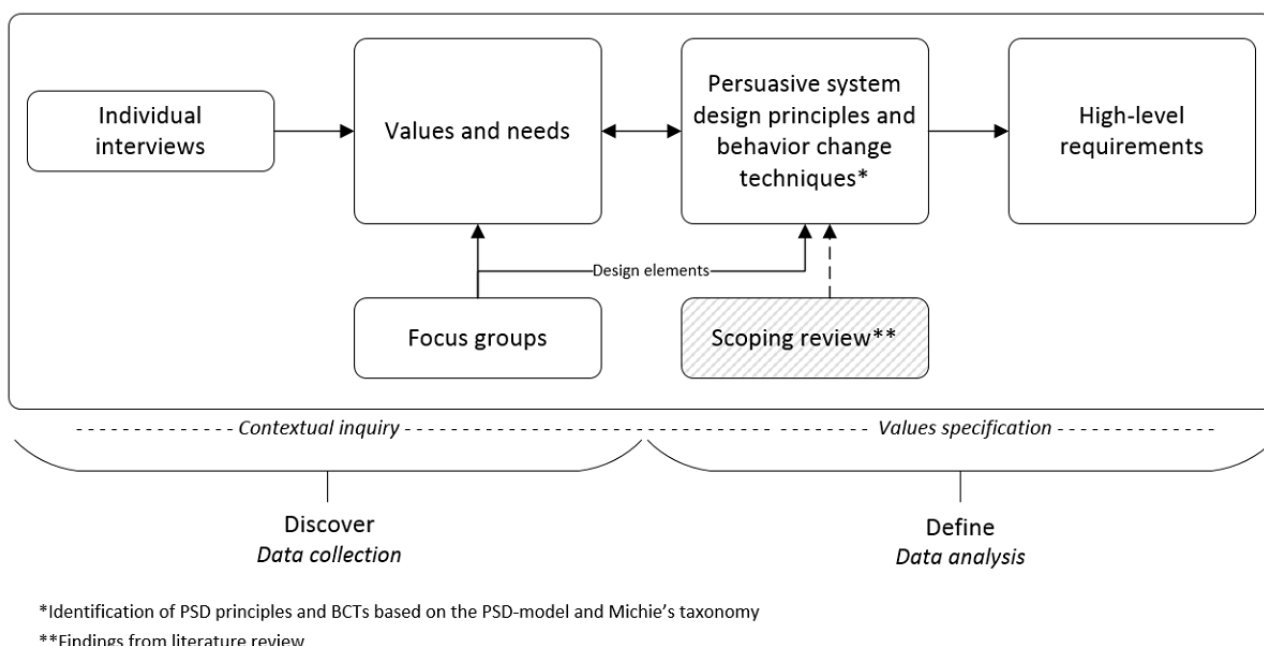
Data Analysis

Semistructured Interviews

As part of the *define* phase, verbatim transcription of audio recordings of the anonymized interviews (n=23) were analyzed by the first author (RA). Thematic analysis was applied, inspired by Brown and Clark [92], using NVivo (QSR International) qualitative analysis software. End user needs were identified through in-depth analysis and inductive coding of the data until saturation was reached [92]. A second researcher (MS) analyzed 17% (4/23) of the transcribed interviews for analysis validation. Overall, 2 researchers (RA and MS) participated in all of the coding of the identified needs into value categories (performed independently of each other), and a third researcher (JW) participated in validating 1/3 of the coding. The findings were continuously discussed within the research team during this process (RA, LN, MS, JW, LP, and JH). All steps contributed to the development of a comprehensive understanding of the data and to identify a representative and specific set of end user values and needs, as expressed by the key stakeholders [75,76,93]. Multimedia Appendix 3 presents examples from the analysis process related to the key values.

Focus Groups

The design elements and related needs identified during the 3 focus group sessions were analyzed, documented, and combined by the first author into one overview in a mind mapping tool (XMind, XMind Ltd) [94]. The findings were discussed with the service designer, digital designer, and the Scrum product owner participating in the focus groups, next to the research team. The identified design elements served as input for the analysis and identification of PSD principles and BCTs to meet end user values and needs (Figure 3).

Identifying PSD Principles, BCTs, and Design Requirements

To identify PSD principles and BCTs, the PSD model by Oinas-Kukkonen [26], behavior change taxonomy by Michie [43], previously conducted scoping review [23], and design elements from the focus group sessions were employed (Figure 3). Through small iterations in the analysis process, needs could be linked to one or more end user values. These could then be linked to possible PSD principles and BCT clusters, leading to the formulation of high-level requirements and suggested design features. The identification analysis was performed by the first author and one of the coauthors (RA and MS). A third coauthor (JW) validated 1/3 of the analysis with regard to inconsistencies and disagreements (randomly selected) and contributed to the requirement development. Data analysis and requirement development were continuously discussed by the research team. This identification process resulted in an overview of PSD principles and BCTs aimed at supporting end user values and needs as well as high-level requirements of an eHealth intervention aiming to support weight loss maintenance. The high-level requirements will allow for innovation during the next steps of the design and development process (Figure 1).

Results

Key Stakeholders and Demographics

Key Stakeholders

The key stakeholders identified were prospective end users (ie, people with obesity aiming to maintain weight after initial weight loss) and health care providers from primary and secondary health care providing health services to the target group. Additional key stakeholders identified were policy makers (ie, from the Norwegian Directorate of eHealth); representatives from a patient organization (ie, the Norwegian Association for People with Obesity); and researchers

conducting research in behavioral medicine, obesity, and weight loss maintenance with clinical experience. See Figure 4 for a final key stakeholder map. In addition, the multidisciplinary design, software, and content development teams involved in the research project were consulted for expert opinions related

to digital design and development (included as indirect stakeholders). Figure 2 presents an overview of key stakeholder involvement in each of the separate research processes in this study (N=50).

Figure 4. Key stakeholder map.
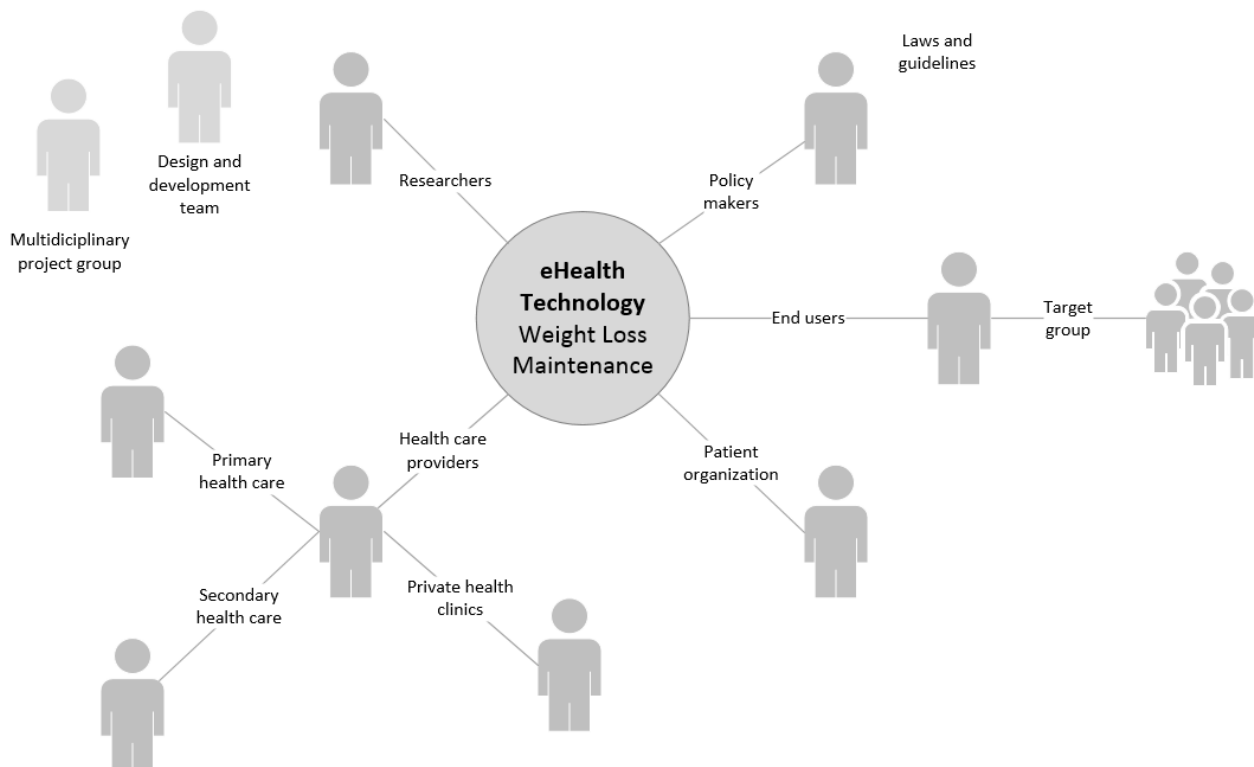
Prospective End Users Involved: General Characteristics

A total of 21 prospective end users, 67% (21/50) women, median age 53 years (range 24-70 years), participated in individual interviews and focus groups. As presented in Multimedia Appendix 4, approximately two-thirds of the participating end users reported having had obesity and weight maintenance challenges for most of their life. All end user participants reported having tried various diets (eg, low-calorie and/or low–calorie diet) and/or conservative treatments (eg, lifestyle programs) to lose weight. They all also described having had several attempts to maintain their weight loss. All participants reported owning a smartphone, tablet, and/or computer and using them on a daily basis for various purposes (eg, social

media, news, weather forecasts, financial services, and buying public transport tickets).

Values and Needs of People Aiming to Maintain Weight After Weight Loss

The results from the individual interviews with end users (n=11) and other key stakeholders (eg, health care personnel, presented in Figure 2; n=13) revealed that end user needs could be clustered into 8 themes (ie, key values) to maintain weight after weight loss. The 8 key values identified included *self-management*, *personalized care*, *motivation*, *feel supported*, *positive self-image*, *health*, *happiness*, and *autonomy*. See Figure 5 for details, including subcategories identified for each of the 8 interconnected values, with some of them overlapping one another.

Figure 5. Overview of identified key end user values.
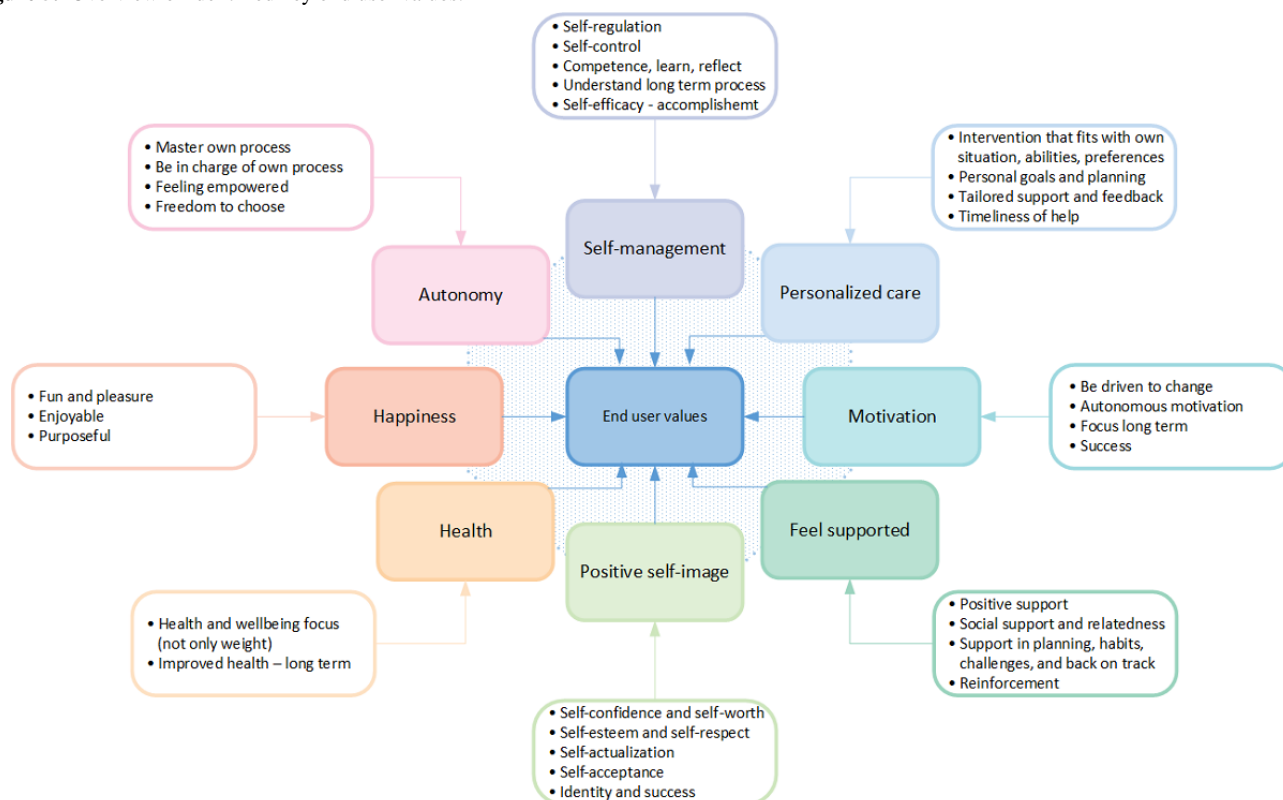
Each specific key value (ie, high-level need or requirement) was generally supported by identified needs. Needs refer to demands and low-level requirements described by end users as aspects to be solved or met to address their weight maintenance problem. In this study, several needs were identified that underlie or support a specific key value (ie, high-level needs). In certain cases, a need could support multiple interrelated key values. For example, the need for *positive feedback on good and bad days to keep up with routines and healthy habits* could be linked to the value *motivation* and to the value *feel supported*. Similarly, the need for *personalized self-monitoring* could be linked to the value *self-management* and the value *personalized care*. Multimedia Appendix 3 provides an overview of examples indicating how *low-level* end user needs were connected to, and embodied, the key values. The following sections describe the findings related to each of the 8 identified key values.

Key Value: Self-Management

Self-management or *self-regulation* was one of the key values highlighted by participants as important to enable sustainable weight maintenance. Many people who struggle to prevent weight regain have difficulties regulating their eating behavior, engaging in physical activities, and establishing new habits:

> *Many lack competence and skills to do what they actually have learned in practice...* [Health care personnel]

In particular, self-regulation strategies and skills to implement into daily routines (eg, planning), dealing with challenges (eg, tempting situations and *back on track* when weight increases), and strategies for coping with emotional eating and impulse control were highlighted as essential needs within the concept of self-management:

> *Weight maintenance is all about planning. I would like to have different plans for different situations, a plan A for "normal days and weeks" and a plan B when "crisis occurs," for example cake at work or holidays, to be able to withstand the constant feeling of hunger and tempting situations.* [End user]

For self-management, comprehension of one's own process to create awareness and reflection around one's own behavior was considered an essential need by end users and other key stakeholders:

> *To understand my own behavior and choices is really important. I can reflect and become aware of my own behavior.* [End user]

> *Help to identify automated behavioral patterns and unhealthy behaviors (eg, eating unhealthy food), self-monitoring of weight and healthy habit support is of essence to understand and regulate behavior...* [Health care personnel]

During the self-management process, the experiences of accomplishment, mastery, and success were deemed especially important to be able to "keep up" over time:

> *I try to focus on what I actually have reached so far. Sometimes I even try old trousers that are now too big...* [End user]

All stakeholders expressed the need for reliable and easily accessible information. As stated by one of the end users, "a lot of information is available for weight loss, but not for weight maintenance."

Key Value: Personalized Care

All stakeholders emphasized that needs and requirements differ depending on the person and whether the focus is on weight loss or weight gain prevention. The differences described included aspects such as knowledge, support, and differences in self-management strategies to regulate eating behaviors, negative thoughts, and/or emotions:

> *People differ, and they experience different challenges over time.* [Health care personnel]

Tailored support to reach goals and establish healthy habits, including small personal steps (ie, subgoals), and to get personal feedback and "information that fits me" (end user) were other needs expressed within the personalized care value:

> *I would like to choose my own habits, goals and sub-goals to work on, such as reaching a weight target, eating more fruit or walk 6000 steps per day, and to keep track of my own progress.* [End user]

Key Value: Motivation

Finding motivation to keep up with the new habits was highlighted by all stakeholders as an important factor to *keep focus* and maintain weight. End users stated that being able to identify individual *motivators* or *reasons* (ie, intrinsic motivation) for maintaining weight and being able to select goals related to individual purpose were essential for staying motivated over time. Other key stakeholders described being able to focus on important purposes, defined by the end users themselves, could contribute to "strengthening of self-regulation capacity and the autonomous motivation required" for the necessary processes of change:

> *The real goal is to find out who you want to be and how you want to live your life, and set your own targets, what's important to you?* [Health care personnel]

Examples of personal *motivators* described by end users included having good health, being more social, being able to play with grandchildren, being able to work, fit into *normal clothes*, and feeling confident. Receiving tailored (social) support and confirmation were also described by end users as crucial to stay motivated during difficult times or when struggling to self-regulate. Other key stakeholders described aspects such as practicing new habits and realistic goal setting, instant reinforcements (ie, rewards, tailored feedback) related to goals and outcomes, reminders of past success, and information about positive versus negative aspects of weight maintenance or regain as being of essence for continued motivation.

Key Value: Feel Supported

Feeling supported, with positive feedback to keep focus and stay motivated to maintain weight, was seen as crucial by all stakeholders. Several end users mentioned that they lacked support in their social surroundings and/or avoided social settings as much as possible. They described that being able to receive confirmation and (social) support, the feeling of not being alone, and having support on *good and bad days* or in *crisis situations* were important needs to be met during weight maintenance:

> *It feels safe when I am part of something and get follow up. It would be nice to have a technology that supports, especially if the weight goes in the wrong direction, in tempting or crisis situations.* [End user]

Other important aspects mentioned by end users included positive reinforcement (eg, praise and self-selected rewards), *just-in-time* support (eg, tailored feedback and 24/7 availability), and the ability to learn from others by sharing experiences with people in similar situations (eg, social learning, social comparison, cooperation, and social facilitation). These statements were supported by other key stakeholders:

> *It is really important that the technology motivates and supports me to keep up...* [End user]

> *They need a motivator, somebody cheering on the sideline, to be able to continue with the healthy habits to maintain weight.* [Health care personnel]

Key Value: Positive Self-Image

To feel confident and believe in oneself was mentioned by most stakeholders as being of essence for weight loss maintenance. The participating end users also stated that the focus should be on the whole person, not solely on their weight:

> *I want to be seen as a person.* [End user]

Several end users and other key stakeholders described that many have low self-confidence after several attempts at weight reduction:

> *Repetition is the clue and regular confirmation, to not fall back into old routines. If they fall back, they lose their confidence even more.* [Health care personnel]

Positive focus on one's strengths and capabilities, the experiences of success, respect for oneself, identity, and self-worth were other areas identified by all end users as important to address to create a more positive self-image and support self-actualization:

> *I feel confident when I know that I do the right things to stay on track.* [End user]

Key Value: Health

Several end users described that the cost of maintaining weight after weight loss tends to outweigh the benefits. End users, therefore, stated that they not only want to focus solely on weight but also on positive health effects of a healthier lifestyle and of being less heavy. They indicated that *having a good health* was one of their main initial motivators to lose weight. Other key stakeholders highlighted aspects such as focusing on short- and long-term health effects of target behavior as important factors to support a feeling of enhanced control and improved health and well-being:

> *We try to focus on their health and well-being—on the positive effects—some still feel the costs more than the benefits.* [Health care personnel]

Key Value: Happiness

The end users expressed a preference to bring attention to enjoyable, fun, and positive benefits of healthy behaviors:

> *Healthy diet and good health should be something fun and nice, not something that gives me a bad conscience and demands a lot of energy.* [End user]

To make physical activities more fun, several end users reported using activity trackers or smartphones to monitor activity levels.

> *Many are highly motivated in the beginning of the weight maintenance phase, until they notice they are still hungry and start to struggle with the same things as in the past. To stay motivated, finding physical activities that bring joy or added purpose and are worth continuing can help.* [Health care personnel]

In terms of technology, special attention was also directed toward fun and enjoyable elements to stimulate happiness and positive feelings, so that a healthy lifestyle could be followed to maintain weight:

> *A technology that gives a fun and meaningful experience...I like to get inspired in a positive and fresh way, maybe with colors, illustrations, symbols and movies.* [End user]

The stakeholders also described being able to focus on short- and long-term health rewards of daily habits, through monitoring and rewarding healthy behaviors or targets, as ways to increase pleasure and happiness. One health care personnel said, "A technology that gives joy and a sense of meaning or added value will be used."

Suggestions and information to make healthy eating more attractive and tastier were also described as ways to increase pleasure and happiness. Other key stakeholders also described factors such as making healthy eating more attractive and enjoyable, being able to get information on the health benefits of being active, and to see body composition improve as important factors for gaining insight.

Key Value: Autonomy

Many end users described the freedom to choose and be in charge as essential aspects to independently master their own behavior change process. They stated that mastery could be reached by having knowledge, abilities, and skills to achieve a healthy lifestyle and continue to do so. The end users also expressed that having customized plans (eg, for weekdays, weekends, holidays) would help induce a feeling of free choice, mastery, and control. Other key stakeholders also supported this:

> *I would like several modules, not dependent on each other...menus to choose from. A personal choice of what to register, which data to share, and when.* [End user]

> *I think the user should be able to choose their own functions and focus...some need more focus on diet, while other need structuring and planning, like a calendar or reminders.* [Health care personnel]

All stakeholders described the following factors as important to be able to choose and continue with healthy habits: facilitating a better understanding of the end users' own process and behavior, enhancing the understanding of how the body and mind work, and learning about health and strategies for self-regulation:

> *It is important to realize that you have a choice. It is not the surroundings that decide what you eat or do.* [Health care personnel]

Identified PSD Principles, BCTs, and Design Requirements of an eHealth Weight Loss Maintenance Intervention

Focus Group Results: Design Elements

Focus group findings provided an overview of potential *design elements* that could meet end user needs when aiming to maintain weight after weight loss. See Table 1 for details. The findings indicate that some design elements were considered more important to incorporate in eHealth weight loss maintenance interventions than others. However, most design elements were characterized as *must have* during focus groups 2 and 3.

Table 1. Focus group results with identified design elements.

Design elements (cards)	Focus group 1 (end users)	Focus group 2 (end users)	Focus group 3 (health care personnel)
Goal setting	++[a]	++	++
Planning	++	++	++
Motivating messages	++	++	++
Personalization	++	++	++
Self-monitoring	++	++	++
Visualization	++	++	++
Tailoring	++	++	++
Feedback	++	++	++
Knowledge[b]	++	++	++
Decision support[b]	++	++	++
Suggestions	++	+[c]	++
Rewards	+	++	++
Reminders	+	++	++
Wearables or sensor technology	+	++	++
Practice habits	+	++	++
Rehearse on situations or challenges	+	++	++
Social support: health care personnel	+	++	++
Social support: other users	+	+	++
Social support: virtual coach	+	_[d]	+
Gamification elements	−	+	+
Metaphors	−	−	−
Avatar	−	−	−

[a]Elements characterized as *must have.*

[b]Design elements suggested by focus group participants.

[c]Elements characterized as *nice to have.*

[d]Elements characterized as *not needed.*

Needs identified during open brainstorming in the focus groups were linked to the design elements. See Table 2 for details. In the analysis process examining the results, the 8 key values were linked to the needs identified.

Table 2. Focus group results with design elements to support weight maintenance needs.

V1[c]	V2[d]	V3[e]	V4[f]	V5[g]	V6[h]	V7[i]	V8[j]	Needs	Design elements cards (PSD[a] and/or BCT[b])
✓[k]	_[l]	—	—	—	✓	✓	✓	• Set own goals (personalized) • Subgoals (eg, eat more fruit)	• Goal setting: • BCT—goals and planning
—	—	—	—	—	—	✓	✓	• Build strategies and plans for different situations (eg, weekdays, weekends, holidays, and birthdays) • Crisis plan and support plan (eg, on "bad days" or periods) • Back-on-track strategy (eg, when drawback occurs, weight increases)	• Planning: • BCT—goals and planning

✓	✓	—	—	—	✓	—	—	• Positive or motivational messages:
								  • Personalized feedback when reaching goals or achievements (eg, self-selected rewards, motivational words, and reminders of personal drivers for losing weight)
								  • On a "difficult" day or period
								  • When using the app (eg, "Welcome back! What's the status?")
								  • Show positive health effects
								  • No negative feedback
									• Motivating messages:
									  • PSD—praise
									  • BCT—feedback and monitoring

✓	—	—	—	—	✓	✓	—	• Personal goals, monitoring, plans, and reminders • Personal messages (praise) and self-selected rewards • Several modules, interdependent (personal choice)	• Personalization: • PSD—personalization

—	—	—	—	—	—	—	✓	• Long-term monitoring of behavior, goals, and plans (eg, through visualizations)
								• My diary (easy and quick self-selected registrations)
									• Self-monitoring:
									  • PSD—self-monitoring
									  • BCT—feedback and monitoring

—	—	—	—	—	✓	—	✓	• Understand own behavior: holistic insights health and well-being (eg, weight, activity, emotion, sleep, and stress)	• Visualization: • BCT—feedback and monitoring
✓	—	—	—	—	—	—	—	• Smart, tailored feedback related to individual lifestyle, goals, and behavior (eg, automatic activity trackers) • Automatic adaptation: favorite modules and interests easily available	• Tailoring: • PSD—tailoring

Key values								Needs	Design elements cards (PSD[a] and/or BCT[b])
V1[c]	V2[d]	V3[e]	V4[f]	V5[g]	V6[h]	V7[i]	V8[j]		

—	✓	✓	—	—	✓	—	✓	• Health promoting (eg, positive spin-offs of maintaining weight and healthy behavior)
								• Positive "boost" messages related to behaviors or habits
								• Positive feedback on "bad days" (eg, personal values, positive goal focus, and earlier achievements)
								• Tailored support (eg, related to crisis plan, tips and feedback, and timely)
								• Smart, tailored feedback-based automatic monitoring and registrations
									• Feedback:
									  • PSD—praise
									  • BCT—feedback and monitoring

—	—	—	—	—	✓	✓	✓	• Competence and skills to regulate behavior, emotions, and thoughts
								• Information:
								  • How the body works after weight loss and recent research and knowledge to maintain weight
								  • Effects and benefits of healthy behaviors and maintaining weight (eg, fun facts)
								  • Preferably presented through a choice of text or sound, supported by visuals and movies
									• Knowledge:
									  • BCT—shaping knowledge and natural consequences

—	✓	✓	—	—	—	—	✓	• Reflect on behaviors and decisions
								• Support in making healthy choices
									• Decision support:
									  • PSD—reduction and suggestions
									  • BCT—antecedents, goals and planning, (eg, action planning), and associations

—	—	—	✓	—	✓	—	✓	• The best evidence-based strategies and tips to maintain weight
								• Focus on health and well-being (eg, how to reach goals and keep up with healthy habits)
								• Practical tips (eg, self-regulation, when technology is not enough)
									• Suggestions:
									  • PSD—suggestions
									  • BCT—regulation

—	—	—	—	✓	✓	—	—	• Rewards (eg, points, trophies) related to goals and targets (eg, weight and activity goals) • Self-selected rewards for motivation	• Rewards: • PSD—rewards
—	—	—	—	—	—	—	✓		

XSL•FO
RenderX

| Key values | | | | | | | | Needs | Design elements cards (PSD[a] and/or BCT[b]) |
V1[c]	V2[d]	V3[e]	V4[f]	V5[g]	V6[h]	V7[i]	V8[j]		

								• Reminders:
								  • My goals, values, and plan
								  • Healthy habits to maintain weight and how the body works
								  • When it goes well and when "off track" (eg, reminders of past successes, trouser that is too big, before and after pictures)
									• Reminders:
									  • PSD—reminders and simulation
									  • BCT—associations (ie, prompts and cues), self-belief and identity

—	—	—	—	✓	✓	—	✓	• Automatic registrations and automatic behavior trackers (eg, weight, activity trackers, wearables, sensors, and smart devices)
								• Ease of use, easy monitoring, and long-term storage of data
								• Personal contact or helper (eg, family or friend for motivation and support)
									• Wearables or sensor technology:
									  • PSD—self-monitoring
									  • BCT—feedback and monitoring (ie, biofeedback)

—	—	✓	—	—	—	✓	✓	• Practice (new) healthy habits
								• Keep up with daily routines and healthy habits in the long term
									• Practice habits:
									  • PSD—reduction and rehearsal
									  • BCT—repetition and substitution (ie, habit formation)

—	—	✓	—	—	—	✓	✓	• Train and prepare for risk situations or tempting situations (eg, "what-if plans," impulse control, and self-regulation)	• Rehearse on situations or challenges:
									  • PSD—rehearsal
									  • BCT—repetition and substitution and goals and planning

—	✓	—	—	—	✓	—	✓	• Contact with coach or health care personnel or general practitioner (eg, when technology is not enough)
								• Support or personal helper (eg, family, friend, or other users or peers through social forum or chat or inspirational user stories) to share experience, learn about health-related behaviors from others, and cooperate
									• Social support:
									  • PSD—social learning, social cooperation, and social facilitation
									  • BCT—social support and comparison of behavior

—	—	—	—	✓	✓		—	• Points and trophies when reaching goals and targets to keep focus and motivation
								• Animated coach for motivation and joy
									• Gamification elements:
									  • PSD—rewards
									  • BCT—reward and threat

[a]PSD: persuasive system design.

[b]BCT: behavior change technique.

[c]V1: personalized care.

[d]V2: feel supported.

[e]V3: positive self-image.

[f]V4: health.

[g]V5: happiness.

[h]V6: motivation.

[i]V7: autonomy.

[j]V8: self-management.

[k]The value was highlighted by the focus groups based on the needs and related design elements identified.

[l]The value was not highlighted by the focus groups based on the needs and related design elements identified.

The findings indicate that several design elements, supported by PSD principles and BCTs, can be applied to facilitate and motivate short- and long-term behavior changes. Positive *feedback, associations* (eg, prompts and cues), and rewards (eg, earning points and receive trophies) were identified as motivational elements to inculcate new behavior and for short-term goal setting during the focus group sessions. For long-term behavior change, rewarding strategies linked to, for example, *self-selected rewards*, *identity* (eg, reminders of past successes), *natural consequences* (eg, information about health consequences of performing the behavior), and *social support* (eg, social cooperation and social learning) were elements identified as potential motivators for sustainable change.

The findings presented in Tables 2 and 3, supported by findings from the previously mentioned scoping review [23], served as input in the analysis to identify PSD principles, BCTs, and requirements of an eHealth intervention to support end user values and needs.

Table 3. Identified behavior change technique clusters supporting end user values.

Behavior change technique clusters based on behavior change taxonomy by Michie	Key values							
	V1[a]	V2[b]	V3[c]	V4[d]	V5[e]	V6[f]	V7[g]	V8[h]
Scheduled consequences	—[i]	—	—	—	—	—	—	—
Reward and threat[j]	—	—	—	—	✓[k]	✓	—	—
Repetition and substitution[l]	—	✓	—	—	—	—	✓	✓
Antecedents	—	—	—	—	—	✓	✓	✓
Associations[j,l]	—	—	—	—	—	—	✓	✓
Covert learning	—	—	—	—	—	✓	—	—
Natural consequences[m]	—	—	—	✓	—	✓	✓	✓
Feedback and monitoring[j,l,m,n]	✓	✓	✓	✓	✓	✓	✓	✓
Goals and planning[j,l,m,n]	—	✓	—	—	—	✓	✓	✓
Social support[j,l,m,n]	—	✓	✓	—	—	✓	—	✓
Comparison of behavior	—	—	—	—	—	—	—	✓
Self-belief[m]	—	—	✓	—	✓	✓	✓	✓
Comparison of outcomes	—	—	—	—	—	✓	—	—
Identity[m]	—	—	✓	—	✓	✓	✓	—
Shaping knowledge[l,m,n]	✓	—	—	✓	—	—	✓	✓
Regulations[m]	—	—	✓	—	✓	✓	—	✓

[a]V1: personalized care.

[b]V2: feel supported.

[c]V3: positive self-image.

[d]V4: health.

[e]V5: happiness.

[f]V6: motivation.

[g]V7: autonomy.

[h]V8: self-management.

[i]Not identified BCTs supporting key values.

[j]Behavior change techniques (BCTs) mentioned applied to stimulate motivation and/or adherence in weight loss maintenance interventions (for long-term change), identified in the previously performed scoping review [23].

[k]Identified BCTs supporting the key values.

[l]Most frequently applied behavior change techniques in weight loss maintenance interventions, identified in the previously performed scoping review [23].

[m]The most frequently identified behavior change techniques in relation to key values in this study.

[n]BCTs included in eHealth interventions that found significant effects for weight loss maintenance, identified in the previously performed scoping review [23].

Identified PSD Principles, BCTs, and Design Requirements of an eHealth Intervention to Support End User Values and Needs

The end user values and needs identified were mapped using BCT taxonomy by Michie [43] to identify BCT clusters of potential relevance for eHealth interventions supporting long-term weight loss maintenance. See Table 3 for the results of this mapping and analysis.

Of the 16 BCT clusters, 15 could be linked to the identified 8 key values for eHealth weight loss maintenance interventions.

Most BCTs were linked to the key values *self-management* (11/16), *motivation* (11/16), and *autonomy* (9/16). BCT clusters supporting 4 key values or more included *natural consequences* (eg, information about health and emotional consequences), *feedback and monitoring* (eg, related to weight and behavior), *self-belief* (eg, focus on past success and positive self-talk to raise self-confidence), *goals and planning* (eg, goal setting, action planning, and problem solving, including relapse prevention and coping planning of risks for weight regain), *identity* (eg, focus on personal strengths or purpose for behavior change associated with the new behavior), *shaping knowledge* (eg, advice and strategies related to diet, physical activity, and

XSL•FO
RenderX

behavior change), *regulations* (eg, skills to regulate or reduce negative emotions), and *social support* to maintain healthy behaviors.

As indicated in Table 3, *goals and planning*, *feedback and monitoring*, and *social support* were identified as important for weight loss maintenance support by the analysis of BCTs to meet end user values as well as findings from the previous scoping review [23]. The BCT clusters *self-belief*, *natural consequences*, and *identity* were also highlighted as being of essence to meet end user values. However, in the previous scoping review, these BCT clusters were not recognized as frequently applied techniques [23]. The identified BCT clusters

in Table 3 (including the underlying techniques) were considered during the high-level requirements specification to illustrate possible combinations of BCTs and PSD principles. Multimedia Appendix 3 presents examples of the BCTs in relation to values, needs, and high-level requirements.

To map the PSD principles relating to end user values and needs, the PSD model [26] was applied. See Table 4 for details. The findings indicate that most of the identified PSD principles were from the *primary task* (6/7) and *dialog support* (6/7) categories of the PSD model [26], followed by the *social support* (4/7) and *credibility support* (4/7) categories.

Table 4. Identified persuasive system design principles supporting end user values.

Persuasive system design principles	Key values							
	V1[a]	V2[b]	V3[c]	V4[d]	V5[e]	V6[f]	V7[g]	V8[h]
Primary task support								
Reduction	✓[i]	—[j]	—	—	—	—	✓	✓
Tunneling	—	—	—	—	—	—	—	—
Tailoring[k,l,m]	✓	✓	✓	✓	✓	✓	✓	✓
Personalization[k,l,m]	✓	✓	✓	✓	✓	✓	✓	✓
Self-monitoring[k,l,m,n]	✓	—	—	✓	✓	✓	✓	✓
Simulation[l]	✓	—	—	—	—	✓	—	✓
Rehearsal	—	✓	—	—	—	—	✓	✓
Dialog support								
Praise[k,m]	✓	✓	✓	✓	✓	✓	—	✓
Rewards[m]	✓	—	—	—	✓	✓	—	—
Reminders[l]	✓	—	—	—	—	✓	—	✓
Suggestions[k]	—	✓	—	✓	—	✓	✓	✓
Similarity	—	—	—	—	—	—	—	—
Liking	—	—	—	—	✓	—	—	—
Social role	—	✓	—	—	—	✓	—	—
System credibility support								
Trustworthiness	—	—	—	✓	—	—	—	✓
Expertise	—	—	—	✓	—	—	—	—
Surface credibility	—	—	—	—	—	—	—	—
Real-world feel	—	✓	—	—	—	—	—	—
Authority	—	—	—	—	—	—	—	—
Third-party endorsements	—	—	—	—	—	—	—	—
Verifiability	—	—	—	✓	—	—	—	—
Social support								
Social learning	—	—	—	—	—	✓	—	✓
Social comparison	—	—	—	—	—	✓	—	✓
Normative influence	—	—	—	—	—	—	—	—
Social facilitation	—	—	—	—	—	—	—	✓
Cooperation	—	✓	—	—	—	—	—	—
Competition	—	—	—	—	—	—	—	—
Recognition	—	—	—	—	—	—	—	—
Others[o]								
Feedback[k,l,m,n]	✓	✓	✓	✓	✓	✓	—	✓
Goal setting[k,l,m,n]	✓	—	—	—	—	✓	✓	✓
Social support[k,m,n]	—	✓	✓	—	—	✓	—	✓

[a]V1: personalized care.

[b]V2: feel supported.

[c]V3: positive self-image.

XSL•FO
RenderX

PSD principles supporting 4 or more key values included *personalization* (eg, of goals, self-selected rewards, and content), *tailoring* (eg, of feedback and information), *praise* (eg, positive, motivating feedback to keep up with healthy behavior), *self-monitoring* (eg, of behavior that supports reaching goals), *feedback* (eg, tailored in relation to already accomplished goals, outcomes, and behavior), *suggestions* (eg, about healthy eating habits), *goal setting* (eg, of outcome and target behavior), and *social support* (eg, through social learning, social comparison, cooperation, or social facilitation by humans or a virtual coach and/or peers). Multimedia Appendix 3 presents examples of the PSD principles identified and how they can be combined with BCTs when specifying high-level requirements for persuasive eHealth technologies supporting weight loss maintenance.

The key values presented under the *Values and Needs of People Aiming to Maintain Weight After Weight Loss* section and in Tables 3 and 4 represent the wants of the end users and their reasons for those wants. The identified BCT clusters and PSD principles indicate techniques that can be applied to change behavior to maintain weight. Many of the BCT clusters, including underlying techniques, were identified as important to address long-term behavior change (eg, *habit formation*). The requirement specification included in Multimedia Appendix 3 shows examples of how BCT clusters, such as *repetition and substitution,* can be combined with PSD principles such as *suggestions* and *reduction*, to support sustained behavior change through eHealth technologies.

Discussion

Principal Findings

This study aimed to provide insight into the design of eHealth interventions supporting behavior change for long-term weight loss maintenance. The goal was to identify values and needs of people with obesity aiming at maintaining weight after weight loss, and to identify PSD principles, BCTs, and design requirements that meet these values and needs in eHealth interventions. The results revealed specific interconnected key end user values (ie, main drivers of behavior or high-level requirements) and needs (ie, demands or low-level requirements), and also provided insight into BCT clusters and PSD principles and how they can be combined to support end user needs in eHealth interventions.

End User Values and Needs to Maintain Weight After Weight Loss

The findings from the semistructured interviews and focus groups with key stakeholders were consistent and complemented each other with respect to the 8 interconnected key values that were identified in this study: *self-management, personalized care, motivation, feel supported, positive self-image, health, happiness, and autonomy.* The key values (ie, main drivers of behavior or high-level requirements) were each supported by identified needs (ie, demands or low-level requirements), sometimes supporting multiple key values. The identified values were interconnected, rather than strictly separated, and dynamic, as the drivers and individual needs can change during long-term maintenance of lost weight.

The 8 identified key values may drive the need for behavior change that is required to successfully maintain weight over time, and motivation and chance for success may increase as these values are met. This means that in eHealth interventions aiming to facilitate sustainable weight loss maintenance, supporting the identified key values in this study could be of essence.

Several of the overlapping end user values identified in this study are in line with research and theories arguing that *most behavior is multi-motivated* and that a value-driven and positive approach to health and well-being is required to achieve sustainable behavior change [95-98]. In such a holistic approach, behavior tends not to be determined by one value or specific goal but by several dynamic needs simultaneously [96,97,99,100]. Theories about motivation, self-regulation, and social learning [101-104] are likely to be of use in the design of eHealth interventions emphasizing values and needs, as identified in this study. For example, the self-determination theory [103,105] offers a macro theory of human motivation and personality, suggesting that motivation to change is dependent on whether people's needs for competence, autonomy, and social relations or relatedness are met. A such, self-determination theory could potentially act as a framework to support several of the identified values and needs, including the individual needs for knowledge and skills (ie, competence) to self-manage, and individual differences with regard to personality, psychological, and/or motivational profile [103,105-108].

PSD Principles, BCT Clusters, and Design Requirements to Meet End User Values and Needs

Several PSD principles and BCT clusters were identified as potentially promising to meet the identified key end user values. Findings from this study are consistent with the results from the scoping review previously conducted by the research team [23], indicating that *goals and planning*, *feedback and monitoring*, and *social support* should be pursued long term to maintain weight [22,23,107,109-114]. However, this study revealed some BCT clusters not previously identified (eg, *identity* and *self-belief*) [23] that could be important to maintain long-term weight loss.

The challenge related to weight loss maintenance for people with obesity is evident, as many people struggle to maintain weight and may easily fall back into old habits following weight loss [6]. This can be explained by the nature of habits, as habits are often learned, automatic, and sometimes even unconscious processes formed through repetition [25]. In this study, BCTs within the *repetition and substitution* clusters (eg, *habit formation* and *habit reversal*) were identified. *Habit formation* can be integrated into eHealth interventions to support the formation of new and lasting habits and might be a missing link in the effort to address the challenges related to weight loss maintenance [7,10,14,15,22,115-121]. To adapt to the dynamic changes in individual needs over time and to support contextual changes, *personalization* and *tailoring* were identified PSD principles of essence in this study.

Findings from this study point to a combination of several PSD principles and BCTs as a necessity to meet end user values and needs in eHealth weight loss maintenance interventions. Persuasive technologies and application of PSD principles in digital behavioral interventions might be key to support long-term weight loss maintenance, especially when combined with BCTs and behavior change theories [23,47]. However, evidence and guidance on how to combine these strategies are nonexistent [23]. PSD principles can be applied to create engaging, user-friendly, persuasive technologies that can support healthy choices and decisions and may be particularly important in *the adoption of new habits* (eg, *suggestions*, *reduction, self-monitoring, rewards,* and *feedback*). On the other hand, BCTs seem to be important for sustainable behavior change, and the identified techniques (eg, *goals and planning*, *monitoring and feedback*, *repetition and substitution*, *associations*, *antecedents, regulation of emotions*, *self-belief*, and *identity*) can be learned and put into daily life, even without technology.

In this study, for some of the key values, only a few supporting BCT clusters and/or PSD principles could be identified. This could indicate that some BCTs or PSD principles have yet to be discovered, possibly because some needs may be future or latent needs. It could also indicate that BCT clusters (eg, *covert learning*) [43] that are not fully desired through identified end user values or needs, but could none the less be important for the required behavior change, should receive particular attention in technology design and may need to be combined with PSD principles (eg, *rehearsal*) [26] to support the user and create the aspired impact on behavior change.

Implications for Future Design and Practice

There are numerous potential advantages of digital health technology. Digital technologies can, for example, overcome time and place barriers, with improved accessibility (24/7 availability) and timely support; enable people to track and aggregate real-time data from sensors and smart devices; and may also support personalized and tailored interventions, depending on user behaviors and needs [122]. To date, however, research examining the design for sustainable behavior change is still sparse [23,123].

If centered around end user needs and priorities, digital self-management support systems have the potential to meet several of the key values identified in this study [104,124-128], which might lead to the long-term behavior change needed to maintain weight. The success of such systems also depends on the incorporation of facilitating factors linked to weight maintenance, including self-regulation strategies (eg, cognitive and emotion regulation, goal setting, effective coping, and problem-solving skills), and increased confidence in own ability to self-regulate [13,104,126-128].

Reversal of established behaviors and habits can be challenging, particularly over time [129,130]. However, applying a combination of the identified BCTs and PSD principles (eg, *repetition and substitution, identity, feedback and monitoring, goals and planning,* and *personalization*) in future eHealth interventions, linking healthy habits to self-determined goals and identity (eg, individual purpose, health or life goal), might contribute to sustainable behavior change and long-term weight control [52,105,108,125,129,131-136]. According to end users and other key stakeholders, many experience that it costs more than it benefits to change behaviors for continued weight control after weight loss. As identified in this study, digital technologies that are designed to support personal and autonomous motivation, self-efficacy, self-regulation skills, positive body image, self-selected rewards, and health and well-being (eg, not only weight) could also contribute to long-term weight control and improve the *cost-benefit* ratio that many people experience while trying to balance behaviors to prevent weight regain [23,103,104,111,125,131,134,137,138].

Although eHealth technologies can contribute to improve health care, support sustainable behaviors promoting health and well-being [139], and increase impact and uptake through personalized and tailored design [36,45,122,140-142], eHealth interventions can only be effective if actually used. This means that persuasive and engaging technologies, as examined in this study, have the potential to support people in engagement and adherence to healthy lifestyles in pursuit of their goal of weight maintenance [30,34].

Behavior change interventions are usually complex, and some studies show how interventions incorporating multiple BCTs, rather than just a few BCTs, tend to have larger effects promoting health behavior change [43,54,143-145]. Digital behavioral obesity interventions that combine PSD principles with behavior change theories have also been shown to produce statistically significant weight loss results more frequently [47]. Specifying one *ideal* combination of PSD principles and BCTs to support sustainable behavior change and weight control in

eHealth interventions might not be possible. Varying individual values and needs, which requires a personalized and tailored approach, make specifying an *ideal* combination at the very least challenging [15,142,146,147]. Investigating which PSD principles, BCTs, and content that is most effective for sustained engagement, for whom and in which context, should be a goal for future research developing or examining digital weight loss maintenance interventions [107].

With several biological, behavioral, psychological, social, and environmental factors interacting, the challenges related to obesity and weight loss maintenance are complex and numerous [16,21]. In this study as well as in the scoping review previously published by the research team [23], personalization and tailoring, including tailored support and personal feedback, were highlighted as factors of great importance for the success of digital interventions targeting behavior change and weight loss maintenance. Individual differences, including personality and psychological profiles that may impact behavior and behavior change, should also be examined and taken into consideration when designing and tailoring such digital interventions in the future [106]. Tailoring based on individual characteristics (eg, personality, motivational orientation) could be of essence when aiming to create effective technological features supporting behavior change [142]. Cognition, affect, and behavior related to individual context have also been identified as core components of engagement and should be considered in the design and development, including the form of delivery and content, of digital interventions aiming to enhance and sustain engagement and support health behavior change [107,147]. In the development process, this issue could be addressed by identifying a variety of user types (eg, personas) [148], and the combination of applied PSD principles and BCTs may potentially be adjusted or *tailored* per user type [142,149]. Ideally, BCTs and PSD principles applied in eHealth behavior change interventions should allow for flexible interaction and adaptability based on the dynamic and changing individual end user needs over time.

Rapid technology developments have changed the way modern organizations approach innovation to create value for end users and stakeholders [68,69,91,150-152]. This study integrates design thinking with BCTs and PSD principles and can guide co-design processes to identify design elements to meet end user values and needs for behavior change in future interventions [64,65,153]. The proposed method used in this study also unlocks and explores current, latent, and future user needs and possible solutions through its iterative and participatory process [21,39,40,68-72,77,154]. The key values, BCTs, and PSD principles identified in this study can provide input to future design and development processes of eHealth weight maintenance interventions. Co-design and prioritization of BCTs and PSD principles with key stakeholders, including designers and developers, may further develop and specify eHealth interventions, allowing for functional requirements and physical design to be adapted and optimized during prototype development and user testing [40,73,91,155]. In addition, multidisciplinary design teams and quick experiments during the entire design and development process may increase the chance for the technology design to successfully meet user needs. However, as shown in this study, some values and needs could be conflicting. The findings from this study also suggest that some needs, PSD principles, or BCTs for sustainable behavior change are yet to be discovered [23,36].

Recommendation for Future Design, Research, and Implementation

Considering the significant individual and public health challenges of obesity and weight maintenance after weight loss, novel, human-centered, and evidence-based solutions aiming to close the gap between weight loss and maintenance of the new weight for long term are needed. First, future design and development of research-based eHealth interventions targeting weight maintenance should aim to investigate how to meet the key values identified in this study and capture expectations and uncovered, or potentially latent, end user needs to support weight maintenance over time [91]. Second, personalization appears to be an important ingredient in successful behavior change. Future research should, therefore, examine the role of personalized and emerging eHealth technologies, supporting the integration of identity-oriented approaches with habit formation and self-regulation strategies to achieve long-term behavior change needed to maintain weight after loss. Third, design elements that facilitate autonomous motivation, create a positive experience and engage users over time, contribute to making healthy behaviors enjoyable, and improve self-efficacy and positive self-image through technology should also be examined [141,147,156].

People, technology, and context are intertwined [39,42,157]. Both services and technologies are shaped through their use, which is why design that allows for interaction and continuous adaptation to user requirements, through personalization and tailoring, is particularly important for adoption and long-term use. To be effective, future solutions aiming to support weight loss maintenance should therefore aim to fit both end user values and needs as well as intervention features and context [42,158]. For researchers, designers, and developers concerned with supporting long-term weight loss maintenance, the findings indicate potential PSD principles and BCTs that can be applied in digital design for sustained behavior change to address the identified key values. Future research should therefore investigate and evaluate how to combine the PSD principles and BCTs in the best way to develop effective, persuasive eHealth interventions supporting sustainable behavior change and weight loss maintenance. Evaluation of technology use also has the potential to reveal user profiles and promising PSD principles and BCTs, addressing the complexity of behavior change to meet individual needs to maintain weight over time [60,76,159]. The value proposition and technology features of future eHealth interventions can be further explored, prioritized, and specified through co-design, prototyping, and testing with key stakeholders, contributing to the next steps of technology design, development, and implementation.

Strengths and Limitations

This study had some limitations. First, a few of the participating end users were still aiming to reduce weight, although their initial weight loss goal was achieved at the point of recruitment. However, this was the case for a few participants only, and as

these participants were also focusing on weight gain prevention, the main results were not considered to be significantly affected. End user needs did differ during weight loss compared with weight maintenance, particularly with regard to dietary monitoring and control (ie, calorie count and registration). In reality, these aspects probably vary and interact. This study therefore likely reflects real-life aspects of weight loss or weight maintenance.

Second, more women than men participated in the study, which may have limited generalizability for the male population. Several other key stakeholders were male, potentially enhancing a representative perspective related to male needs and values when aiming to change behavior and maintain weight after weight loss, with the support of eHealth technology.

Third, not all participating end users had experience with successful long-term weight loss maintenance. Needs following a longer period of weight loss maintenance might be different from the needs following a short-term period. As most end users in this study were recruited within 2 to 3 months following weight loss, this could have influenced the findings in this study with regard to long-term needs. By including end users with positive as well as negative experiences from short- and long-term weight gain prevention, a broad view and understanding of the weight maintenance phenomena could be explored.

Fourth, the participants who volunteered for this study might have been more engaged, motivated, and interested in the use of eHealth technologies than the average weight maintenance key stakeholder. Given the nature of the qualitative methods applied in this study, the findings allow for detailed information about a complex issue. However, more research examining aspects of values and needs to maintain weight loss over time is required to enhance generalizability. Finally, identification of PSD principles and BCT clusters might have been prone to subjectivity by the researchers. However, to prevent subjective interpretation, 3 researchers participated in the validation of identified PSD principles and BCTs to reveal and address inconsistencies in this study.

This study also had several strengths. First, the research and development team were multidisciplinary, with clinical and research expertise within behavioral medicine, obesity, eHealth, and persuasive technology, and included user representatives as well as software developers and designers. Second, the methods applied required high levels of end user and other key stakeholder involvement. Such an approach can be time consuming but might be critical for success when designing and developing eHealth interventions [36]. Third, by applying a variety of qualitative methods and involving various stakeholders [36,75,93], a broad understanding of the issues at hand was achieved, capturing and verifying end user needs from various perspectives (eg, patients, health care personnel, and policy makers). Fourth, the mixed and converging research and design methodologies applied [39,68,72], involving multiple researchers in the data collection and analysis processes, aimed to increase validity and reduce possible researcher bias. Finally, the presentation of values and needs and a thorough description of the analysis process identifying and translating PSD principles and BCTs into high-level requirements aimed at contributing to transparency, understanding, and reproducibility of the findings.

Conclusions

eHealth interventions have the potential to support the regulation of behaviors and maintenance of weight after weight loss. This study contributes to a better understanding of the values and needs of people aiming to maintain weight after weight loss. The translation of values and needs into design elements or features supported by PSD principles and BCTs could play an important role in the design of future eHealth interventions that support sustained behavior change and long-term weight maintenance. How PSD principles and BCTs can be combined in the best way to facilitate behavior change to achieve long-term weight loss maintenance remains to be determined.

The methods described in this study can guide the design of digital interventions and services supporting behavior change to meet end user values and needs in the future. To the best of our knowledge, this is the first study to present insights and suggestions for a new approach on how to identify and translate end user values and needs into PSD principles and BCTs when designing eHealth self-management interventions for sustained behavior change and weight loss maintenance.

Conflicts of Interest

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 92 KB - jmir_v22i11e22598_app1.pdf]

Multimedia Appendix 2
Design elements of feature cards.
[PDF File (Adobe PDF File), 294 KB - jmir_v22i11e22598_app2.pdf]

Multimedia Appendix 3
Value specification, persuasive system design principles and behavior change techniques identification, and high-level requirements development for eHealth interventions supporting weight loss maintenance.
[PDF File (Adobe PDF File), 156 KB - jmir_v22i11e22598_app3.pdf]

Multimedia Appendix 4
Infographics to visualize the demographics of included end users.
[PDF File (Adobe PDF File), 106 KB - jmir_v22i11e22598_app4.pdf]

References

1.   Obesity and Overweight - Key Facts. World Health Organization. 2020. URL: http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-07-01]
2.   Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the global burden of disease study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
3.   GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 6;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
4.   Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. Lancet 2011 Aug 27;378(9793):815-825. [doi: 10.1016/S0140-6736(11)60814-3] [Medline: 21872750]
5.   Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
6.   Christiansen T, Bruun JM, Madsen EL, Richelsen B. Weight loss maintenance in severely obese adults after an intensive lifestyle intervention: 2- to 4-year follow-up. Obesity (Silver Spring) 2007 Feb;15(2):413-420 [FREE Full text] [doi: 10.1038/oby.2007.530] [Medline: 17299115]
7.   McGuire M, Wing R, Klem M, Hill J. Behavioral strategies of individuals who have maintained long-term weight losses. Obes Res 1999 Jul;7(4):334-341. [doi: 10.1002/j.1550-8528.1999.tb00416.x] [Medline: 10440589]
8.   Thomas JG, Bond DS, Phelan S, Hill JO, Wing RR. Weight-loss maintenance for 10 years in the national weight control registry. Am J Prev Med 2014 Jan;46(1):17-23. [doi: 10.1016/j.amepre.2013.08.019] [Medline: 24355667]
9.   Paixão C, Dias CM, Jorge R, Carraça EV, Yannakoulia M, de Zwaan M, et al. Successful weight loss maintenance: a systematic review of weight control registries. Obes Rev 2020 May;21(5):e13003. [doi: 10.1111/obr.13003] [Medline: 32048787]
10. Wing R, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Jul;82(1 Suppl):222S-225S. [doi: 10.1093/ajcn/82.1.222S] [Medline: 16002825]
11. Klem M, Wing R, Lang W, McGuire M, Hill J. Does weight loss maintenance become easier over time? Obes Res 2000 Sep;8(6):438-444 [FREE Full text] [doi: 10.1038/oby.2000.54] [Medline: 11011910]
12. The National Weight Control Registry. 1994. URL: http://www.nwcr.ws [accessed 2020-07-01]
13. McKee H, Ntoumanis N, Smith B. Weight maintenance: self-regulatory factors underpinning success and failure. Psychol Health 2013;28(10):1207-1223. [doi: 10.1080/08870446.2013.799162] [Medline: 23767689]
14. Wyatt H, Grunwald G, Mosca C, Klem M, Wing R, Hill J. Long-term weight loss and breakfast in subjects in the national weight control registry. Obes Res 2002 Feb;10(2):78-82 [FREE Full text] [doi: 10.1038/oby.2002.13] [Medline: 11836452]

15. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]

16. Montesi L, El Ghoch M, Brodosi L, Calugi S, Marchesini G, Dalle Grave R. Long-term weight loss maintenance for obesity: a multidisciplinary approach. Diabetes Metab Syndr Obes 2016;9:37-46 [FREE Full text] [doi: 10.2147/DMSO.S89836] [Medline: 27013897]

17. Jensen M, Ryan D, Apovian C, Ard J, Comuzzie A, Donato K, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]

18. Brockmann AN, Eastman A, Ross KM. Frequency and consistency of self-weighing to promote weight-loss maintenance. Obesity (Silver Spring) 2020 Jul;28(7):1215-1218. [doi: 10.1002/oby.22828] [Medline: 32437055]

19. Pedersen S, Sniehotta FF, Sainsbury K, Evans EH, Marques MM, Stubbs RJ, et al. The complexity of self-regulating food intake in weight loss maintenance. A qualitative study among short- and long-term weight loss maintainers. Soc Sci Med 2018 Jul;208:18-24. [doi: 10.1016/j.socscimed.2018.05.016] [Medline: 29753930]

20. Karfopoulou E, Mouliou K, Koutras Y, Yannakoulia M. Behaviours associated with weight loss maintenance and regaining in a Mediterranean population sample. A qualitative study. Clin Obes 2013 Oct;3(5):141-149. [doi: 10.1111/cob.12028] [Medline: 25586629]

21. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]

22. Cleo G, Isenring E, Thomas R, Glasziou P. Could habits hold the key to weight loss maintenance? A narrative review. J Hum Nutr Diet 2017 Oct;30(5):655-664. [doi: 10.1111/jhn.12456] [Medline: 28150402]

23. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]

24. Smith KS, Graybiel AM. Habit formation. Dialogues Clin Neurosci 2016 Apr;18(1):33-43. [doi: 10.31887/DCNS.2016.18.1/ksmith]

25. Fogg B. Tiny Habits. Boston, Massachusetts, USA: Houghton Mifflin Harcourt; 2015:2015.

26. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Info Syst 2009;24:485-500. [doi: 10.17705/1CAIS.02428]

27. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

28. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]

29. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. JMIR Mhealth Uhealth 2018 Apr 10;6(4):e92 [FREE Full text] [doi: 10.2196/mhealth.4478] [Medline: 29636320]

30. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

31. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]

32. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

33. Nilsen P, Roback K, Broström A, Ellström PE. Creatures of habit: accounting for the role of habit in implementation research on clinical behaviour change. Implement Sci 2012 Jun 9;7:53 [FREE Full text] [doi: 10.1186/1748-5908-7-53] [Medline: 22682656]

34. Kelders S, Kip H. Development and Initial Validation of a Scale to Measure Engagement with eHealth Technologies. In: Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, United Kingdom. [doi: 10.1145/3290607.3312917]

35. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to ehealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]

36. Gemert-Pijnen JE, Kelders S, Kip H, Sanderman R. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London, UK: Routledge; 2018.

37. Héroux M, Watt M, McGuire K, Berardi J. A personalized, multi-platform nutrition, exercise, and lifestyle coaching program: a pilot in women. Internet Interv 2017 Mar;7:16-22 [FREE Full text] [doi: 10.1016/j.invent.2016.12.002] [Medline: 30135822]

38. Arnrich B, Mayora O, Bardram J, Tröster G. Pervasive healthcare. Methods Inf Med 2018 Jan 17;49(1):67-73. [doi: 10.3414/me09-02-0044]

39. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]

40. Lentferink A, Polstra L, D'Souza A, Oldenhuis H, Velthuijsen H, van Gemert-Pijnen L. Creating value with eHealth: identification of the value proposition with key stakeholders for the resilience navigator app. BMC Med Inform Decis Mak 2020 Apr 27;20(1):76 [FREE Full text] [doi: 10.1186/s12911-020-1088-1] [Medline: 32340615]

41. Støme LN, Moger T, Kidholm K, Kværner KJ. Early assessment of innovation in a healthcare setting. Int J Technol Assess Health Care 2019 Jan;35(1):17-26. [doi: 10.1017/S0266462318003719] [Medline: 30744712]

42. Cruz-Martínez RR, Wentzel J, Asbjørnsen RA, Noort PD, van Niekerk JM, Sanderman R, et al. Supporting self-management of cardiovascular diseases through remote monitoring technologies: metaethnography review of frameworks, models, and theories used in research and development. J Med Internet Res 2020 May 21;22(5):e16157 [FREE Full text] [doi: 10.2196/16157] [Medline: 32436852]

43. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

44. Oinas-Kukkonen H. Persuasive System Design: state of the art and future directions. In: International Conference on Persuasive Technology. 2009 Presented at: CPT'09; April 4-9, 2009; Claremont, California. [doi: 10.1145/1541948.1541989]

45. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in ehealth interventions combining self-tracking and persuasive ecoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 1;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]

46. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Evaluation of a web-based lifestyle coach designed to maintain a healthy bodyweight. J Telemed Telecare 2010;16(1):3-7. [doi: 10.1258/jtt.2009.001003] [Medline: 20086259]

47. Sittig S, McGowan A, Iyengar S. Extensive review of persuasive system design categories and principles: behavioral obesity interventions. J Med Syst 2020 Jun 5;44(7):128. [doi: 10.1007/s10916-020-01591-w] [Medline: 32500161]

48. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec;2002(December):2. [doi: 10.1145/764008.763957]

49. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]

50. Brindal E, Hendrie GA, Freyne J. Combining persuasive technology with behavioral theory to support weight maintenance through a mobile phone app: protocol for the motimate app. JMIR Res Protoc 2016 Jan 8;5(1):e5 [FREE Full text] [doi: 10.2196/resprot.4664] [Medline: 26747725]

51. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: the relationship between web features and weight loss. Telemed J E Health 2008 Oct;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]

52. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Hum Behav 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]

53. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. Mhealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]

54. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

55. Soltani H, Arden MA, Duxbury AM, Fair FJ. An analysis of behaviour change techniques used in a sample of gestational weight management trials. J Pregnancy 2016;2016:1085916 [FREE Full text] [doi: 10.1155/2016/1085916] [Medline: 27034836]

56. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. Ehealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]

57. Hesse BW, Shneiderman B. eHealth research from the user's perspective. Am J Prev Med 2007 May;32(5 Suppl):S97-103 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.019] [Medline: 17466825]

58. Neve M, Morgan P, Jones P, Collins C. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]

59. Evans EH, Araújo-Soares V, Adamson A, Batterham AM, Brown H, Campbell M, et al. The NULevel trial of a scalable, technology-assisted weight loss maintenance intervention for obese adults after clinically significant weight loss: study protocol for a randomised controlled trial. Trials 2015 Sep 22;16:421 [FREE Full text] [doi: 10.1186/s13063-015-0931-7] [Medline: 26395774]

60. Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. Obesity (Silver Spring) 2020 Mar;28(3):502-509. [doi: 10.1002/oby.22715] [Medline: 31965748]

61. Kelders SM, van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]

62. Mooc Ehealth: Combining Psychology, Technology and Health - Ehealth: What Are Values? The Open University. University of Twente. 2016. URL: http://www.futurelearn.com/courses/ehealth [accessed 2020-07-01]

63. Stickdorn MS. This is Service Design Thinking: Basics, Tools, Cases. Amsterdam, Netherlands: Wiley; 2014.

64. The Double Diamond: a Universally Accepted Depiction of the Design Process. British Design Council. 2005. URL: http://www.designcouncil.org.uk/news-opinion/double-diamond-universally-accepted-depiction-design-process [accessed 2020-07-01]

65. What is the Framework for Innovation? Design Council's Evolved Double Diamond. British Design Council. 2019. URL: http://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond [accessed 2020-07-01]

66. StimuLabs Metode: Trippel Diamant. DOGA. 2020. URL: http://doga.no/aktiviteter/design-og-innovasjon/innovasjon-i-offentlig-sektor/stimulab/arbeidsmetode/ [accessed 2020-06-22]

67. The Triple Diamond. DOGA. 2020. URL: https://doga.no/en/activities/design-og-innovasjon/stimulation-programme-for-innovation-and-service-design2/hvordan-apne-for-innovasjon/ [accessed 2020-10-23]

68. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]

69. Kolko J. Design Thinking Comes of Age. Harvard Business Review. 2015. URL: http://hbr.org/2015/09/design-thinking-comes-of-age [accessed 2020-01-07]

70. Polaine A, Reason B, Løvlie L. Service Design - From Insight to Implementation. New York, USA: Rosenfeld Media; 2013.

71. Ferreira F, Song E, Gomes H, Garcia E, Ferreira L. New mindset in scientific method in the health field: design thinking. Clinics (Sao Paulo) 2015 Dec;70(12):770-772. [doi: 10.6061/clinics/2015(12)01] [Medline: 26735214]

72. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]

73. Stickdorn MS, Schneider J. This Is Service Design Thinking: Basics, Tools, Cases. Hoboken, New Jersey: Wiley; 2012.

74. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. New York, USA: HarperBusiness; 2009.

75. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999 Dec;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]

76. van Velsen L, Wentzel J, van Gemert-Pijnen JE. Designing ehealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]

77. van Woezik AF, Braakman-Jansen LM, Kulyk O, Siemons L, van Gemert-Pijnen JE. Tackling wicked problems in infection prevention and control: a guideline for co-creation with stakeholders. Antimicrob Resist Infect Control 2016;5:20 [FREE Full text] [doi: 10.1186/s13756-016-0119-2] [Medline: 27213040]

78. van Limburg AH. Implementing Antibiotic Stewardship: Involving Stakeholders in Ehealth. Enschede, The Netherlands: University of Twente; 2016.

79. What is Overweight and Obesity. World Health Organization. URL: https://www.who.int/health-topics/obesity#tab=tab_1 [accessed 2020-07-01]

80. Green J, Thorogood N. Qualitative Methods for Health Research. Newbury Park, California: SAGE Publications; 2018.

81. Gubrium JF, Holstein JA. Focus group interviewing. In: Handbook of Interview Research: Context and Method. Thousand Oaks, California, USA: Sage Publications; 2002.

82. Pope C, Mays N. Qualitative Research in Health Care. Hoboken, New Jersey: Blackwell Publishing Ltd; 2006.

83. Morville PR, Rosenfeld L. Information Architecture for the World Wide Web. Newton, Massachusetts, USA: O'Reilly Media; 2007.

84. Wentzel J, van Velsen L, van Limburg M, de Jong N, Karreman J, Hendrix R, et al. Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med Inform Decis Mak 2014 Jun 5;14:45 [FREE Full text] [doi: 10.1186/1472-6947-14-45] [Medline: 24898694]

85. Simonsen J. In: Robertson T, editor. International Handbook of Participatory Design. Abingdon, England, UK: Routledge; 2013.

86. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mhealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]

87. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KM, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of epio. J Med Internet Res 2020 Jan 21;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]

88. Spook JE, Paulussen T, Paulissen R, Visschedijk G, Kok G, van Empelen P. Design rationale behind the serious self-regulation game intervention 'balance it': overweight prevention among secondary vocational education students in the Netherlands. Games Health J 2015 Oct;4(5):387-400. [doi: 10.1089/g4h.2014.0142] [Medline: 26181575]

89. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

90. DeSmet A, van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]

91. Casanova P. Agile Processes: A Unifying Approach for the Future of Projects. In: PMI Global Congress Proceedings. 2013 Presented at: PGC'13; April 24, 2013; Istanbul, Turkey URL: https://www.researchgate.net/publication/273683645_Agile_Processes_A_Unifying_Approach_for_the_Future_of_Projects

92. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

93. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]

94. XMind - Mind Mapping Software.: XMind URL: http://www.xmind.net [accessed 2020-10-23]

95. Huber M, van Vliet M, Giezenberg M, Winkens B, Heerkens Y, Dagnelie PC, et al. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: a mixed methods study. BMJ Open 2016 Jan 12;6(1):e010091. [doi: 10.1136/bmjopen-2015-010091] [Medline: 26758267]

96. Maslow A. Motivation and Personality. Manhattan, New York, USA: Harper; 1954.

97. Maslow AH. Motivation and personality. In: Frager R, Fadiman J, McReynolds C, editors. Harper and Row. London, England: Longman; 1987.

98. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? Br Med J 2011 Jul 26;343:d4163. [doi: 10.1136/bmj.d4163] [Medline: 21791490]

99. Maslow A. Theory of human motivation. Psychol Rev 1943;50:370-396.

100. Deckers L. Motivation: Biological, Psychological, and Environmental. New York, USA: Routledge; 2018.

101. Parcel G, Kok G, Gottlieb NH. Planning Health Promoting Programs, an Intervention Mapping Approach. Second Edition. San Francisco, USA: Jossey-Bass (Wiley); 2011.

102. Gollwitzer P, Sheeran P. Implementation intentions and goal achievement: a meta‐analysis of effects and processes. Adv Exper Soc Psychol 2006;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]

103. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]

104. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. 1st Edition. New Jersey, US: Prentice-Hall Inc; 1986.

105. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. New York, USA: Springer; 1985.

106. Sutin AR, Ferrucci L, Zonderman AB, Terracciano A. Personality and obesity across the adult life span. J Pers Soc Psychol 2011 Sep;101(3):579-592 [FREE Full text] [doi: 10.1037/a0024286] [Medline: 21744974]

107. Teixeira PJ, Marques MM. Health behavior change for obesity management. Obes Facts 2017;10(6):666-673 [FREE Full text] [doi: 10.1159/000484933] [Medline: 29237167]

108. Silva MN, Markland D, Minderico CS, Vieira PN, Castro MM, Coutinho SR, et al. A randomized controlled trial to evaluate self-determination theory for exercise adherence and weight control: rationale and intervention description. BMC Public Health 2008 Jul 9;8:234 [FREE Full text] [doi: 10.1186/1471-2458-8-234] [Medline: 18613959]

109. Stubbs RJ, Lavin JH. The challenges of implementing behaviour changes that lead to sustained weight management. Nutr Bull 2013 Feb 7;38(1):5-22. [doi: 10.1111/nbu.12002]

110. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2012 Mar;6(1):7-32. [doi: 10.1080/17437199.2010.513298]

111. Dombrowski SU, Knittle K, Avenell A, Araújo-Soares V, Sniehotta FF. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. Be Med J 2014 May 14;348:g2646. [doi: 10.1136/bmj.g2646] [Medline: 25134100]

112. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]

113. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

114. Sniehotta FF, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul;35(4):565-576. [doi: 10.1002/ejsp.258]

115. Swift DL, Johannsen NM, Lavie CJ, Earnest CP, Church TS. The role of exercise and physical activity in weight loss and maintenance. Prog Cardiovasc Dis 2014;56(4):441-447 [FREE Full text] [doi: 10.1016/j.pcad.2013.09.012] [Medline: 24438736]

116. Catenacci VA, Wyatt HR. The role of physical activity in producing and maintaining weight loss. Nat Clin Pract Endocrinol Metab 2007 Jul;3(7):518-529 [FREE Full text] [doi: 10.1038/ncpendmet0554] [Medline: 17581621]

117. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]

118. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of 'habit-formation' and general practice. Br J Gen Pract 2012 Dec;62(605):664-666. [doi: 10.3399/bjgp12X659466] [Medline: 23211256]

119. van der Weiden A, Benjamins J, Gillebaart M, Ybema JF, de Ridder D. How to form good habits? A longitudinal field study on the role of self-control in habit formation. Front Psychol 2020;11:560 [FREE Full text] [doi: 10.3389/fpsyg.2020.00560] [Medline: 32292376]

120. Cleo G, Beller E, Glasziou P, Isenring E, Thomas R. Efficacy of habit-based weight loss interventions: a systematic review and meta-analysis. J Behav Med 2020 Aug;43(4):519-532. [doi: 10.1007/s10865-019-00100-w] [Medline: 31529279]

121. Lally P, van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]

122. Kelders SM, Oinas-Kukkonen H, Oörni A, van Gemert-Pijnen JE. Editorial: special issue on health behavior change support systems. Int J Med Inform 2016 Dec;96:1-2. [doi: 10.1016/j.ijmedinf.2016.07.001] [Medline: 27401468]

123. Niedderer K, Clune S, Ludden G. Design for Behaviour Change: Theories and practices of designing for change, 1st Edition. England, UK: Routledge; 2017.

124. Bergeson SC, Dean JD. A systems approach to patient-centered care. J Am Med Assoc 2006 Dec 20;296(23):2848-2851. [doi: 10.1001/jama.296.23.2848] [Medline: 17179462]

125. Teixeira PJ, Silva MN, Mata J, Palmeira AL, Markland D. Motivation, self-determination, and long-term weight control. Int J Behav Nutr Phys Act 2012 Mar 2;9:22 [FREE Full text] [doi: 10.1186/1479-5868-9-22] [Medline: 22385818]

126. Poulimeneas D, Yannakoulia M, Anastasiou C, Scarmeas N. Weight loss maintenance: have we missed the brain? Brain Sci 2018 Sep 11;8(9):174 [FREE Full text] [doi: 10.3390/brainsci8090174] [Medline: 30208568]

127. Phelan S, Halfman T, Pinto AM, Foster GD. Behavioral and psychological strategies of long-term weight loss maintainers in a widely available weight management program. Obesity (Silver Spring) 2020 Feb;28(2):421-428 [FREE Full text] [doi: 10.1002/oby.22685] [Medline: 31970912]

128. Reyes NR, Oliver TL, Klotz AA, Lagrotte CA, Vander Veur SS, Virus A, et al. Similarities and differences between weight loss maintainers and regainers: a qualitative analysis. J Acad Nutr Diet 2012 Apr;112(4):499-505. [doi: 10.1016/j.jand.2011.11.014] [Medline: 22709701]

129. Ouellette JA, Wood W. Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. Psychol Bull 1998 Jul;124(1):54-74. [doi: 10.1037/0033-2909.124.1.54]

130. Hollands GJ, Marteau TM, Fletcher PC. Non-conscious processes in changing health-related behaviour: a conceptual analysis and framework. Health Psychol Rev 2016 Dec;10(4):381-394 [FREE Full text] [doi: 10.1080/17437199.2015.1138093] [Medline: 26745243]

131. Verplanken B, Sui J. Habit and identity: behavioral, cognitive, affective, and motivational facets of an integrated self. Front Psychol 2019;10:1504 [FREE Full text] [doi: 10.3389/fpsyg.2019.01504] [Medline: 31354563]

132. Elfhag K, Rössner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. Obes Rev 2005 Feb;6(1):67-85. [doi: 10.1111/j.1467-789X.2005.00170.x] [Medline: 15655039]

133. Hooglugt F, Ludden GD. A mobile app adopting an identity focus to promote physical activity (MoveDaily): iterative design study. JMIR Mhealth Uhealth 2020 Jun 15;8(6):e16720 [FREE Full text] [doi: 10.2196/16720] [Medline: 32538795]

134. Kearney MH, O'Sullivan J. Identity shifts as turning points in health behavior change. West J Nurs Res 2003 Mar;25(2):134-152. [doi: 10.1177/0193945902250032] [Medline: 12666640]

135. Gardner B, de Bruijn G, Lally P. A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. Ann Behav Med 2011 Oct;42(2):174-187. [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]

136. Silva MN, Vieira PN, Coutinho SR, Minderico CS, Matos MG, Sardinha LB, et al. Using self-determination theory to promote physical activity and weight control: a randomized controlled trial in women. J Behav Med 2010 Apr;33(2):110-122. [doi: 10.1007/s10865-009-9239-y] [Medline: 20012179]

137. Teixeira PJ, Carraça EV, Marques MM, Rutter H, Oppert J, de Bourdeaudhuij I, et al. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. BMC Med 2015 Apr 16;13:84 [FREE Full text] [doi: 10.1186/s12916-015-0323-6] [Medline: 25907778]

138. Varkevisser RD, van Stralen MM, Kroeze W, Ket JC, Steenhuis IH. Determinants of weight loss maintenance: a systematic review. Obes Rev 2019 Feb;20(2):171-211 [FREE Full text] [doi: 10.1111/obr.12772] [Medline: 30324651]

139. IJsselsteijn W, de Kort Y, Midden C, Eggen B, van den Hoven E. Persuasive Technology for Human Well-Being: Setting the Scene. In: Lecture Notes in Computer Science. 2006 Presented at: PERSUASIVE'06; May 18-19, 2006; Springer, Berlin, Heidelberg.

140. Børøsund E, Ehlers SL, Varsi C, Clark MM, Andrykowski MA, Cvancarova M, et al. Results from a randomized controlled trial testing StressProffen; an application-based stress-management intervention for cancer survivors. Cancer Med 2020 Jun;9(11):3775-3785 [FREE Full text] [doi: 10.1002/cam4.3000] [Medline: 32243717]

141. Niedderer K, Clune S, Ludden G. Design for Behaviour Change for Health and Wellbeing. London, UK: Routledge, Taylor and Francis; 2017.

142. van Velsen L, Broekhuis M, Jansen-Kosterink S, Op den Akker H. Tailoring persuasive electronic health strategies for older adults on the basis of personal motivation: web-based survey study. J Med Internet Res 2019 Sep 6;21(9):11759 [FREE Full text] [doi: 10.2196/11759] [Medline: 31493323]

143. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

144. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]

145. Prochaska JJ, Nigg CR, Spring B, Velicer WF, Prochaska JO. The benefits and challenges of multiple health behavior change in research and in practice. Prev Med 2010;50(1-2):26-29 [FREE Full text] [doi: 10.1016/j.ypmed.2009.11.009] [Medline: 19948184]

146. Fuglestad PT, Rothman AJ, Jeffery RW, Sherwood NE. Regulatory focus, proximity to goal weight, and weight loss maintenance. Am J Health Behav 2015 Sep;39(5):709-720 [FREE Full text] [doi: 10.5993/AJHB.39.5.12] [Medline: 26248180]

147. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to ehealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]

148. Vosbergen S, Mulder-Wiggers J, Lacroix J, Kemps H, Kraaijenhagen R, Jaspers M, et al. Using personas to tailor educational messages to the preferences of coronary heart disease patients. J Biomed Inform 2015 Feb;53:100-112 [FREE Full text] [doi: 10.1016/j.jbi.2014.09.004] [Medline: 25239261]

149. Dam RF, Siang TY. Personas – A Simple Introduction. Interaction Design Foundation. URL: http://www.interaction-design.org/literature/article/personas-why-and-how-you-should-use-them [accessed 2020-09-07]

150. Osterwalder A, Pigneur Y. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, New Jersey: John Wiley and Sons; 2014.

151. Ries E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Australia: Currency; 2011.

152. Osterwalder A, Pigneur Y, Bernarda G, Smith A, Papadakos T. Value Proposition Design: How to Create Products and Services Customers Want. Hoboken, New Jersey: Wiley; Oct 2014.

153. Mirkovic J, Jessen S, Kristjansdottir OB, Krogseth T, Koricho AT, Ruland CM. Developing technology to mobilize personal strengths in people with chronic illness: positive codesign approach. JMIR Form Res 2018 Jun 5;2(1):e10774 [FREE Full text] [doi: 10.2196/10774] [Medline: 30684404]

154. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]

155. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]

156. Petermans A, Cain R, editors. Design for Wellbeing. An Applied Approach. England, UK: Routledge; 2020.

157. Kip H, Sieverink F, van Gemert-Pijnen LJ, Bouman YH, Kelders SM. Integrating people, context, and technology in the implementation of a web-based intervention in forensic mental health care: mixed-methods study. J Med Internet Res 2020 May 26;22(5):e16906 [FREE Full text] [doi: 10.2196/16906] [Medline: 32348285]

158. Everett MR. Diffusion of Innovations. New York, USA: Free Press; 2003.

159. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]

Abbreviations

BCT: behavior change technique
CeHRes roadmap: Center for eHealth Research and Disease Management roadmap
OUH: Oslo University Hospital

PSD: persuasive system design
SH: Sørlandet Hospital
VHT: Vestfold Hospital Trust

XSL•FO
RenderX

Review

Online Patient Recruitment in Clinical Trials: Systematic Review and Meta-Analysis

Mette Brøgger-Mikkelsen[1,2,3], MA; Zarqa Ali[1,2], MD, PhD; John R Zibert[3], MSc, PhD; Anders Daniel Andersen[3], MSc, PhD; Simon Francis Thomsen[1,2], Dr med

[1]Department of Dermatology, Bispebjerg Hospital, Copenhagen, Denmark

[2]Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Studies&Me A/S, LEO Innovation Lab, Copenhagen, Denmark

Corresponding Author:
Mette Brøgger-Mikkelsen, MA
Department of Dermatology
Bispebjerg Hospital
Bispebjerg Bakke 23
Copenhagen, 2400
Denmark
Phone: 45 28901012
Email: mette.mikkelsen@studiesandme.com

Abstract

Background: Recruitment for clinical trials continues to be a challenge, as patient recruitment is the single biggest cause of trial delays. Around 80% of trials fail to meet the initial enrollment target and timeline, and these delays can result in lost revenue of as much as US $8 million per day for drug developing companies.

Objective: This study aimed to conduct a systematic review and meta-analysis examining the effectiveness of online recruitment of participants for clinical trials compared with traditional in-clinic/offline recruitment methods.

Methods: Data on recruitment rates (the average number of patients enrolled in the study per month and per day of active recruitment) and conversion rates (the percentage of participants screened who proceed to enroll into the clinical trial), as well as study characteristics and patient demographics were collected from the included studies. Differences in online and offline recruitment rates and conversion rates were examined using random effects models. Further, a nonparametric paired Wilcoxon test was used for additional analysis on the cost-effectiveness of online patient recruitment. All data analyses were conducted in R language, and $P<.05$ was considered significant.

Results: In total, 3861 articles were screened for inclusion. Of these, 61 studies were included in the review, and 23 of these were further included in the meta-analysis. We found online recruitment to be significantly more effective with respect to the recruitment rate for active days of recruitment, where 100% (7/7) of the studies included had a better online recruitment rate compared with offline recruitment (incidence rate ratio [IRR] 4.17, $P=.04$). When examining the entire recruitment period in months we found that 52% (12/23) of the studies had a better online recruitment rate compared with the offline recruitment rate (IRR 1.11, $P=.71$). For cost-effectiveness, we found that online recruitment had a significantly lower cost per enrollee compared with offline recruitment (US $72 vs US $199, $P=.04$). Finally, we found that 69% (9/13) of studies had significantly better offline conversion rates compared with online conversion rates (risk ratio 0.8, $P=.02$).

Conclusions: Targeting potential participants using online remedies is an effective approach for patient recruitment for clinical research. Online recruitment was both superior in regard to time efficiency and cost-effectiveness compared with offline recruitment. In contrast, offline recruitment outperformed online recruitment with respect to conversion rate.

KEYWORDS

XSL•FO
RenderX

Introduction

Historically, recruitment of participants for clinical trials has been critically dependent upon physician referrals and overall site performance. Increasing needs for more effective recruitment methods have led to "trial and error" models, where a number of different recruitment strategies are utilized and modified according to observed effects on recruitment [1]. Such traditional recruitment strategies include, but are not limited to, soliciting subjects through mail and telephone using health records and registers, media campaigns, newspaper advertisements, and input during radio and television talks [2].

Currently, recruitment for clinical trials continues to be a challenge, as patient recruitment is the single biggest cause of trial delays. Around 80% of trials fail to meet the initial enrollment target and timeline, and these delays can result in lost revenue of as much as US $8 million per day for drug developing companies [3]. As pointed out by Gul and Ali [4], slow and inefficient recruitment may have scientific, economic, and ethical consequences. The costs associated with recruitment are wasted, and data quality is hampered by a reduction in statistical power due to underrecruitment. Recruiting appropriate participants in a sufficient number to fulfill sample size requirements is critical for the validity of the research findings, and failure may lead to invalid or inconclusive results. Further, for traditional offline recruitment strategies, location of trial sites quickly becomes the bottleneck for participant diversity in clinical research, as sites only succeed in recruiting patients within a relatively short radius. Potentially, this results in clinical research that lacks generalizability and makes it difficult for clinical trials to be a cornerstone for providing scientific evidence on the safety and efficacy of novel pharmaceutical compounds.

Using online recruitment strategies, such as social media advertisements, Google search engine advertisements, and other website campaigns, may enable researchers to target specific study populations by demographic characteristics, location, and keywords previously used in potential participants' user profiles. In 2018, Akers et al [5] found that Facebook advertisements gave "flexibility to monitor and modify advertisement tactics based on feedback," and Shere et al [6] argued that social media recruitment should be redefined as an *active* recruitment tool rather than a low-budget *passive* tool, as targeting specific populations effectively yielded high recruitment rates. Similar arguments were made by Watson et al [7], Jones et al [8], and Carter-Harris et al [9] who all reported social media advertisements to be a "viable tool for more efficient and cost-effective recruitment." The potential reach of online recruitment by far exceeds the reach that traditional recruitment methods are able to generate, but whether online recruitment strategies outperform traditional offline recruitment strategies still remains unclear in the literature. Herein, traditional in-clinic recruitment methods are referred to as offline recruitment.

The aim of this study was to evaluate the effectiveness of online patient recruitment by systematically reviewing studies that utilize online strategies for patient recruitment and by conducting meta-analyses comparing online and offline recruitment strategies on the following two recruitment metrics: recruitment rate (the number of patients enrolled in the study on average per month and per day of active recruitment) and conversion rate (the percentage of participants screened who proceed to enroll into the clinical trial). Further, this study investigated the cost-effectiveness of online recruitment compared with offline recruitment in clinical research. In the study, we investigated the following three hypotheses: (1) The recruitment rate is higher in online recruitment compared with offline recruitment; (2) The conversion rate is higher in offline recruitment compared with online recruitment; and (3) The cost per enrolled subject is lower in online recruitment compared with offline recruitment.

Methods

Literature Search

We conducted a systematic review and meta-analysis in accordance with the PRISMA statement [10] searching the following databases: PubMed, EMBASE, Cochrane Library, Web of Science, and Google Scholar. The search was conducted between February and May 2020, and was carried out by combining keywords within the following three topic domains: "online/remote/web-based," "patient/participant/subject recruitment," and "clinical trial/study." Two reviewers (MBM and ZA) independently screened all titles and abstracts, and any discrepancies were resolved through discussion. Apart from duplicates, studies were excluded based on the content of the abstract if there was no clear indication that they investigated the feasibility of online patient recruitment.

Screening and Study Selection

Studies were included in the systematic review provided that they fulfilled the following inclusion criteria: (1) Clinical studies using online recruitment and/or prescreening of patients for randomized controlled trials (RCTs), observational studies, and online surveys relevant to the focus of this study and (2) Clinical studies using a fully virtual approach from screening to data collection. For the meta-analysis, only studies that compared online patient recruitment with offline patient recruitment were included. Studies were excluded if they fulfilled one or more of the following exclusion criteria: (1) non-English papers; (2) systematic reviews; and (3) other (nonrelevant online programs, eg, parenting training programs).

Data Extraction

For data collection, the first reviewer (MBM) used a structured form to extract the following qualitative and quantitative data: (1) study design and year; (2) study location by country; (3) total number of participants enrolled in the study; (4) online and offline recruitment metrics (full recruitment period in months, number of days with active recruitment, number of patients completing prescreening, number of patients enrolled by recruitment method, and costs); and (5) age and gender of the participants in the study. For studies where stated data collection was inadequate, the study authors were contacted and necessary data were obtained when possible.

XSL•FO
RenderX

Outcomes Assessed

The primary outcome variable for analysis was an aggregate measure of recruitment effectiveness defined by the recruitment rate and conversion rate. Two analyses were carried out. First, the recruitment rate was defined as the number of patients recruited per month for the entire recruitment period, and second, the recruitment rate was defined as the number of patients recruited per day of active recruitment days. As online campaigns are mostly run in shorter periods with active online advertisements, removing the days in between advertisements was expected to provide a more realistic understanding of the recruitment. For offline recruitment, days in between nonactive recruitment were removed in the second analysis. The conversion rate was defined as the percentage of patients screened who proceeded to enroll into the clinical trial. Prescreening is either an online prescreening questionnaire, or on-ground screening or telephone call for assessing primary eligibility. If necessary, we recalculated the metrics needed for the analyses where possible.

The secondary outcome of interest was the cost-effectiveness of online recruitment compared with offline recruitment. To standardize cost data, we adjusted all costs to US$ using XE Live Exchange Rate 2020.

Statistical Methods

We pooled effect sizes based on the Mantel-Haenszel method for both recruitment rate (incidence rate ratio [IRR]) and conversion rate (risk ratio [RR]). A random effects DerSimonian and Laird meta-analysis [11] was used to report both incidence rates and relative risks. Further, we calculated heterogeneity (DerSimonian and Laird estimator), which was examined using the I^2 statistics, and 95% prediction intervals were calculated. The Hartung-Knapp adjustment for a random effects model was used to calculate 95% CIs, reflecting the uncertainty in heterogeneity. All data analyses were conducted in R (R Foundation for Statistical Computing) [12] (packages included devtools, meta, dmetar, and pbkrtest), and $P<.05$ was considered significant.

Meta-Analysis

The only criterion for carrying out the meta-analysis was the availability of sufficient outcomes. We considered that any amount of statistical heterogeneity would be acceptable, as studies included in this paper recruited for a wide range of therapeutic areas and trial interventions. Hence, we expected high heterogeneity among the included studies. However, the recruitment strategy was far more homogeneous between studies, and therefore, we considered the findings worth reporting. We performed two separate meta-analyses to determine the effectiveness of online patient recruitment.

Recruitment Rate

The effect size for recruitment rate was calculated as an incidence rate (IR), as this rate signifies how many events occur within a standardized timeframe. Effect sizes were pooled to generate the IRR, examining the relation between the incidence rate in the online recruitment group (IR_{online}) and the one in the offline recruitment group ($IR_{offline}$). The pooled data for recruitment rate was presented in a forest plot showcasing whether the recruitment rate was in favor of online or offline recruitment.

Conversion Rate

The conversion rate was calculated as an event rate with relative risk as the effect size, since event rate data deal with the number of persons experiencing an event in each group and the total sample size in each group. Effect sizes were pooled to generate RRs as the summary measure. As for recruitment rate, the pooled data were presented in a forest plot showcasing whether the conversion rate was in favor of online or offline recruitment.

Additional Analysis

This study presents a cost-effectiveness analysis for studies included in the meta-analyses defined as cost per enrollee. As the distribution in our two paired data sets was nonnormal, cost-effectiveness was examined by a nonparametric paired sample Wilcoxon test that does not assume a specific underlying distribution of data.

Risk of Bias

The risk of bias was assessed using the Cochrane Risk of Bias Tool for randomized studies and the ROBINS-I tool for nonrandomized studies. The studies were assessed for risk of bias in relation to our review question and not the study authors' research question.

Results

Literature Search

A total of 3861 articles, including references from articles, were identified for possible inclusion based on their titles, and of these, 395 were selected for abstract screening. Of these, 135 studies were selected for full-article review, and 61 studies that reported the use of online patient recruitment without meeting any of the exclusion criteria were included in the systematic review. Of the selected studies, 23 studies investigated the feasibility of online patient recruitment compared with offline patient recruitment and were included in the meta-analysis after removing duplicates (Figure 1).

XSL•FO

RenderX

Figure 1. Flow diagram of the literature search.
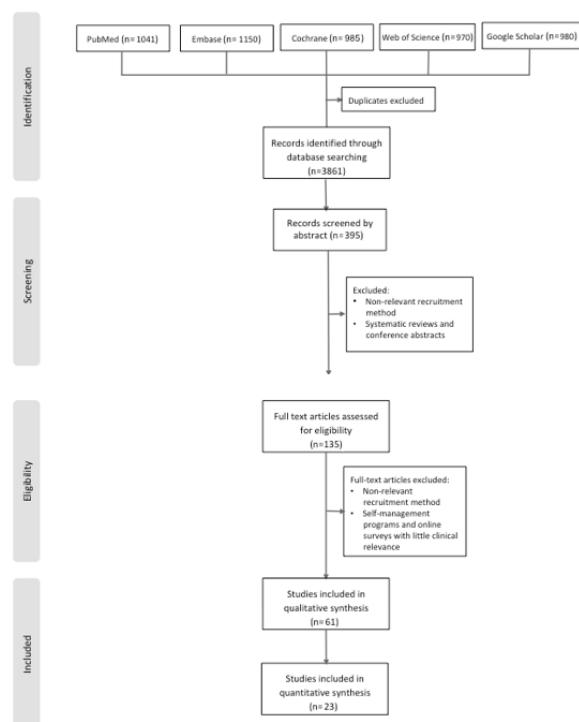
Characteristics of All Included Studies

Multimedia Appendix 1 shows the detailed characteristics of all studies included in this review. Most were conducted in the United States or Australia, and together, the studies covered a wide range of therapeutic areas and trial interventions, with the majority of studies recruiting either adult smokers [7,9,13-24], men who have sex with men [8,25-33], or pregnant and postpartum women [6,34-39]. Of all the studies included, 15 reported a full clinical trial utilizing online recruitment itself (eg, [40-44]), while 46 of the studies described the recruitment strategy utilized in a clinical trial reported in a separate paper. In total, 39 studies covered the recruitment strategy for RCTs, 14 studies covered the recruitment strategy for observational research trials, and eight studies covered the recruitment strategy for online surveys.

Recruitment Strategy

Of the 61 studies included in this review, 42 studies concluded that targeting potential study participants online was an effective tool for recruitment. Only four studies reported that online recruitment was not effective [17,45-47], which might reflect the timing and time period of the online strategy in these four studies. The remaining studies did not conclude on effectiveness. A total of 55 studies used paid advertisements in their online recruitment strategy, of which 42 studies specifically used Facebook. In 2016, Adam et al [36] used Facebook advertisements to target pregnant women for an RCT and found that online strategies compared with offline strategies recruited a representative population and that recruitment rates had been "dramatically improved." Similar findings were reported by Cowie et al [48], who found that Facebook yielded highly efficient and cost-effective results when targeting people aged 60 years or older. In contrast, Rait et al [17] found that Facebook advertisements expanded reach when recruiting young adults

for a smoking cessation trial; however, only a small proportion was eligible for the study, and offline methods were therefore superior to online methods both in regard to cost and time-efficiency. In total, 26 studies further used websites relevant for the specific trial, for example, popular drug control websites when recruiting for a drug abuse prevention study [34,49,50].

Demographics

On examining the overall patient demographics of the included studies, most papers recruited participants aged 18 years or above, and the vast majority of studies recruited both men and women. Furthermore, 46 of the studies targeted a so called *hard-to-reach* population, reaching a sample that has previously been shown to be difficult to recruit owing to either stigmatization, such as smoking cessation research (eg, [7,21]) and HIV prevention research (eg, [29,30,51-53]), or underrepresentation and low prevalence, such as research involving children with fetal alcohol spectrum disease [54] and some psychiatric conditions [55-59]. In 2014, Shere et al [6] successfully utilized social media to recruit a hard-to-reach population of women in the periconceptional period for an RCT, and among others, Morgan et al [60] effectively recruited a sample for an RCT evaluating a depression intervention through several online strategies [61].

Characteristics of the Studies Included in the Meta-Analyses

For the meta-analyses, we included 23 studies, and all reported data on comparing the effectiveness of online and offline recruitment strategies. In total, 14 of the studies included in the meta-analyses recruited for RCTs reported in a separate paper [6-8,19-21,27,30,36,45,46,62-64]. Of the 23 studies, approximately 25% began online recruitment at a later time point than offline recruitment owing to low recruitment rates

through offline recruitment approaches [6,28,36,45,62]. Data extracted from the 23 studies were used to calculate the recruitment rate, conversion rate, and cost per enrollee. Multimedia Appendix 2 shows the aggregated characteristics of the studies included in all the meta-analyses.

Effectiveness of Online Patient Recruitment

Principal Findings

This study found that online recruitment strategies are superior to offline recruitment initiatives when measuring recruitment effectiveness by recruitment rate and cost-effectiveness. With online strategies, participants are recruited faster and more cost-effectively. However, this study found that offline recruitment outperforms online recruitment when converting potential participants to actual enrollees, which altogether is in line with our study hypotheses.

For recruitment rate, we found online recruitment to be significantly more effective when examining the recruitment rate for active days of recruitment, where 100% (7/7) of the studies included had a better online recruitment rate compared with the offline recruitment rate (*P*=.04). When examining the entire recruitment period in months, we found that 12 of 23

studies (52%) had a better online recruitment rate compared with the offline recruitment rate. For the conversion rate, we found that only 4 out of the 13 studies (31%) had a better online conversion rate compared with the offline conversion rate.

Meta-Analyses

Recruitment Rate

For the seven studies included in our first meta-analysis, we found that the recruitment rate for online recruitment was superior to that for traditional offline recruitment when comparing the number of active recruitment days presented as an IRR. For online recruitment, this corresponded to the number of days the advertisements were active on social media or other relevant websites. For offline recruitment, this was the number of days with active on-ground recruitment and active days of advertisements in newspapers, on busses, etc. Pooling the data from the seven studies, we found that online recruitment strategies recruited subjects significantly faster than offline recruitment strategies (IRR 4.17, 95% CI 1.12-15.59, *P*=.04). Hereby, online recruitment strategies yielded 4.17 times more participants per day of active recruitment compared with offline recruitment strategies. As expected, we found high heterogeneity between studies (I^2=100%) (Figure 2).

Figure 2. Recruitment rate for active days of enrollment for online and offline recruitment. Value <1, in favor of offline recruitment; value >1, in favor of online recruitment.

Study	Online Events	Online Time	Offline Events	Offline Time	Incidence Rate Ratio	IRR	95%-CI	Weight
Jones et al	154	205	94	205		1.64	[1.27; 2.12]	14.3%
Bracken et al	16	330	29	1470		2.46	[1.33; 4.52]	14.0%
Adam et al	25	26	45	210		4.49	[2.75; 7.32]	14.1%
Partridge et al	68	45	173	510		4.45	[3.37; 5.90]	14.3%
Bull et al	1079	270	697	360		2.06	[1.88; 2.27]	14.4%
Raymond et al	1764	56	134	378		88.86	[74.55; 105.92]	14.4%
Parsons et al	220	35	694	162		1.47	[1.26; 1.71]	14.4%
Overall effect						**4.17**	**[1.12; 15.59]**	**100.0%**

Heterogeneity: I^2 = 100% [100%; 100%], *p* = 0

0.01   0.1   1   10   100

For our second meta-analysis, all 23 articles were included. Here, we compared the full period of recruitment for both online and offline strategies, looking at the number of months of recruitment from the start of the first advertisement or campaign to the end of the last advertisement or campaign. Pooling the data from the 23 included studies, we found that online

recruitment was similar to offline recruitment with respect to effectiveness when nonactive days within the recruitment period were also included in the analysis (IRR 1.11, 95% CI 0.62-1.97, *P*=0.7). Again, we found high heterogeneity between studies (I^2=99%) (Figure 3).

Figure 3. Recruitment rate for the entire period of recruitment online and offline in months. Value <1, in favor of offline recruitment; value >1, in favor of online recruitment.
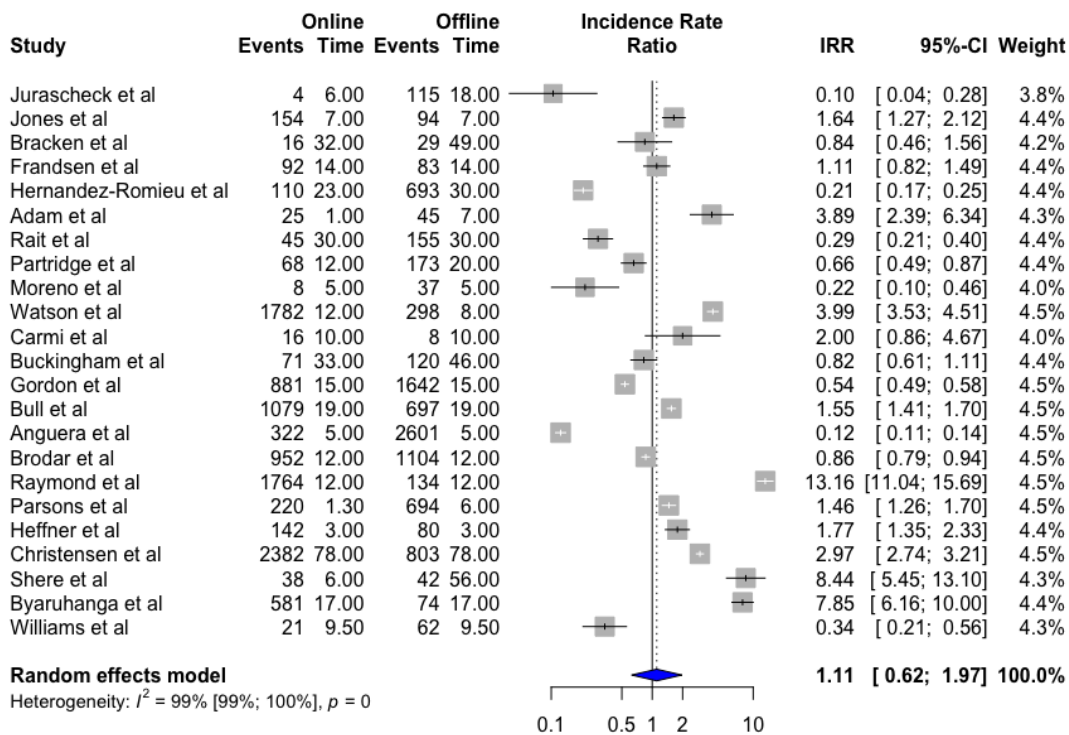
Conversion Rate

For our meta-analysis on conversion rate, we included 13 of the 23 articles. We found that traditional offline recruitment strategies were superior to online recruitment strategies when comparing the percentage of participants screened who proceed to enroll into the clinical trial presented as a RR. For online recruitment, potential participants were screened through online questionnaires when clicking on online advertisements. For offline recruitment, potential participants who had shown interest in participating in the clinical trial were screened on ground or through telephone calls. Pooling the data from the 13 included studies, we found that online recruitment strategies converted significantly fewer potential participants into enrolled subjects compared with offline recruitment strategies (RR 0.8, 95% CI 0.67-0.96, $P=.02$). Hereby, offline recruitment strategies are more effective in converting participants who are screened into enrolled subjects in clinical trials. As for the recruitment rate, we found high heterogeneity between studies ($I^2=96\%$) (Figure 4).

Figure 4. Conversion rate for online and offline recruitment. Value <1, in favor of offline recruitment; value >1, in favor of online recruitment.
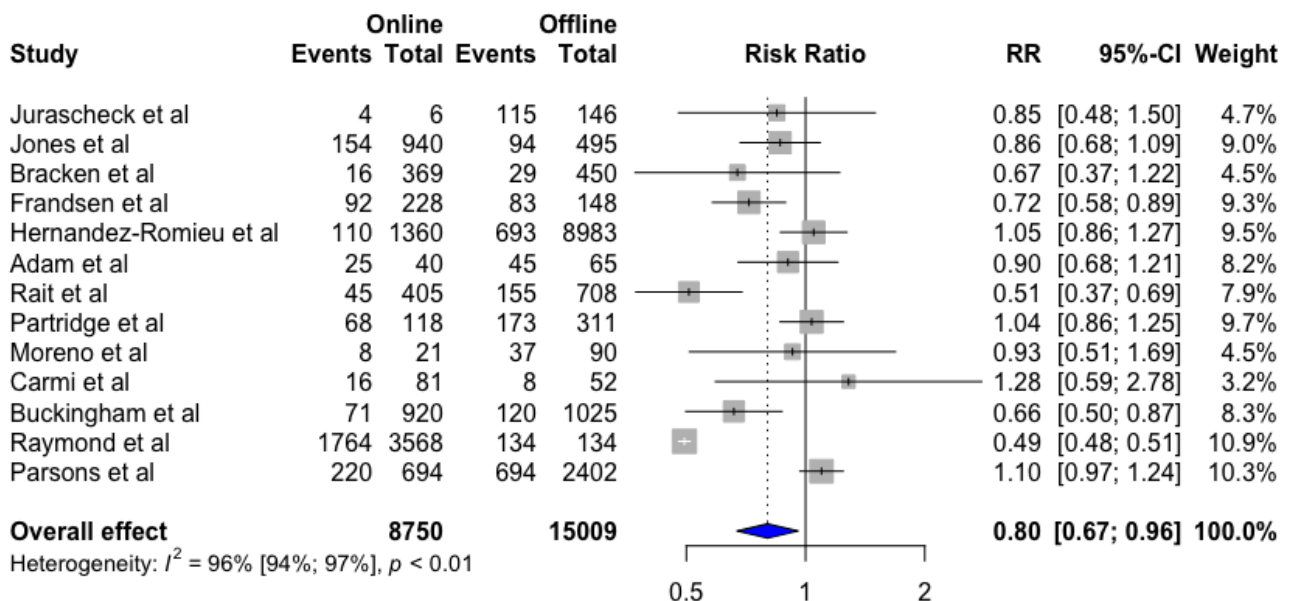
Additional Analysis

Cost-Effectiveness of Online Recruitment

Of the 23 articles included for the meta-analyses, 13 studies reported data on the cost per enrolled participant [7,8,17,19-21,24,28,46,47,57,62,63]. The median cost per enrollee for online recruitment strategies was US $72 (range US $3.9-251.2), while the median cost per enrollee for offline recruitment strategies was US $199 (range US $19.1-839.0). However, the average cost per enrolled participant varied between the different studies, and in total, 4 out of the 13 included studies reported online recruitment to be less cost-effective compared with offline recruitment. We found a significant difference between online and offline cost-effectiveness (*P*=.048), with a V value of 17, meaning that there was a large difference between the two groups. Multimedia Appendix 3 presents cost per enrollee for all studies included in this analysis.

Risk of Bias for Studies Included in the Meta-Analyses

For the 14 randomized studies, none of the studies were deemed to be at risk of bias owing to allocation concealment, blinding, and missing outcome data when assessing the articles in relation to the focus of this review. Hence, no performance bias, detection bias, or attrition bias was identified. For the nine nonrandomized studies, no bias related to confounding, selection of participants, and classification of interventions was identified. However, for the studies included, the representativeness of the recruited samples was often discussed. In total, 13 of the studies included tested differences in representativeness relative to samples obtained through offline recruitment [6-8,19-21,24,28-31,36,47]. Of these, only 31% (4/13) reported that samples recruited online were similarly representative to samples recruited offline, whereas 69% (9/13) found relevant differences among the two groups. Characteristics that were most frequently reported to be imbalanced included gender (no consistent trend found), age (no consistent trend found), and education (higher education overrepresented for offline recruitment). A summary of the risk of bias can be found in Multimedia Appendix 2.

Discussion

Overall Findings

This study found that targeting potential participants using online remedies is an effective approach for patient recruitment for clinical research. Online recruitment strategies were superior in regard to time efficiency and cost-effectiveness compared with offline recruitment strategies. In contrast, offline recruitment strategies outperformed online recruitment strategies when examining conversion rate. To our knowledge, this is the first time a meta-analysis has been performed on the effectiveness of online patient recruitment. Our findings are consistent with findings from previous reviews on online patient recruitment [65].

Quantitative Analysis

The recruitment rate reported in this study was only significantly better for online recruitment when days in between active recruitment were removed from the analysis (*P*=.04). This emphasizes that advertisements and campaigns on social media are efficient in a relatively short time period, after which the recruitment effectiveness starts dropping. Such findings were also reported by Juraschek et al [45], who concluded that offline recruitment was superior to online recruitment in a randomized trial recruiting cancer survivors. Therefore, online strategies should be run in intermittent campaigns to maximize full recruitment capacity, and they could also be of value when considering a short-term high-output recruitment solution. As pointed out previously, a large proportion of the studies that compared online and offline methods utilized online recruitment strategies only after realizing that offline strategies did not provide enough participants [6,7,28,36,46,62]. This could have an impact on the results for online recruitment, because the online solution here is at risk of being a short-sighted sticking-plaster solution. If online targeting had been the primary strategy of recruitment, the results might have been in the favor of online recruitment in this case [66].

Offline conversion rates were found to be significantly higher than online conversion rates (*P*=.02), as originally hypothesized. This could be due to sites having existing health records of suited patients for specific studies. For prescreening in an offline setting, it is only patients who have already shown interest in participating or referrals who already have prequalified for enrollment that are actually screened by inclusion and exclusion criteria. Further, the reach of traditional offline recruitment strategies, such as measurements of how many people read newspapers or how many people listen to campaigns on radio, cannot be quantified to the same extent as the reach for online advertisements. Hence, we do not know the actual conversion rates for specific initiatives, and therefore, the results on the conversion rate in this paper could be overestimated for offline recruitment strategies.

Our cost-effectiveness analysis showed that online recruitment is more cost-effective compared with offline recruitment. For online strategies, organic reach, such as social sharing generated online, may contribute to high cost-effectiveness, as it gives an exponential increase in message exposure, and Shere et al [6] referred to online sampling as "snow-ball recruitment," which involves letting social media automatically expand the reach of a similar population. Even so, the cost differences arguably can be even larger, since the costs associated with offline recruitment strategies may have been underestimated. This is because for offline recruitment, the costs of personnel time were only included in very few of the cost measurements in the articles. On the other hand, setting up online advertisements and tracking and monitoring them demands personnel hours, especially in cases where the staff is inexperienced in online advertisements and campaigns and needs thorough training. As alluded earlier, to obtain the best out of online recruitment efforts, trained personnel specifically dedicated to plan proper recruitment strategies and mitigations are required. As Facebook and other online sites are in a constant state of flux, best practices for recruitment strategies here are hard to develop since no fit-for-all model works [5].

Qualitative Analysis

This review found that the majority of published studies that used online patient recruitment were recruiting a hard-to-reach population and utilizing the potential for online recruitment tools to specifically target an underrepresented population. For this finding, it is worth noting that online remedies also have the advantage of no in-person contact, as these hard-to-reach populations often are rather stigmatized (eg, drug abusers). Further, studies included in this review targeted men and women aged 18 years or above, with no studies reporting age as a critical bottleneck for online recruitment. Overall, this study found that online recruitment was favorable compared with offline recruitment. For clinical researchers, these findings could be of great value when designing future research studies.

Shortcomings of Online Patient Recruitment

Online methods have recruited samples with atypical demographic characteristics compared with offline methods, suggesting that the internet may reach a different population of subjects compared with samples recruited offline and hereby introduce a risk of recruitment bias. For instance, one study recruited smokers who were more likely to be nondaily smokers, exhibit high motivation to quit, and use alcohol than reported in other studies [13,67], and Bull et al [27] recruited a sample through Facebook, where the vast majority of individuals were Caucasian. On the other hand, this could introduce higher diversity in the trial, reflecting real-world demographics. Therefore, some studies argue that a combination of online and offline strategies to reach sample targets yields the most representative and unbiased samples [47]. However, during recent years, being on the internet has become the norm for almost everyone, and differences in populations may therefore be declining. Furthermore, design of advertisements is also likely an important factor, as different designs and persona character types may influence the audience signing up for a study.

Strengths and Limitations

The vast majority of studies included in both the review and meta-analyses were RCTs, which are considered the "gold standard" for clinical trials. Further, online recruitment can be easily monitored, and data metrics from online sites are detailed and easy to obtain. However, although online recruitment metrics are widely available, many studies included in our meta-analysis lacked recruitment data for both online and offline recruitment methods. As such, the data were not complete for all of the 23 studies included in the meta-analysis. In addition, our 95% prediction intervals should be interpreted with caution because prediction intervals have been reported to be less reliable in meta-analyses with unbalanced study sizes [68]. Although speculative, more complete data could potentially have reduced any ambiguity in the results. Finally, what we have considered for our review may not be exhaustive, as there could be underlying factors that have not been investigated. We collected a limited amount of demographic data from the articles, and only touched upon the discussion related to demographics and the two recruitment methods.

Implications for Clinical Trials

Inability to meet recruitment targets in clinical research is the biggest cause of trial delays. Our findings suggest that online strategies for patient recruitment for clinical trials can speed up recruitment and potentially reduce trial delays, as this paper substantiates that online recruitment is both more time-efficient and cost-effective compared with offline recruitment. However, this paper also suggests that to obtain the best possible results, both effort and money should be invested in online recruitment campaigns. To maximize the recruitment rate, online strategies should not be seen as an add-on to offline recruitment, but should be a primary recruitment strategy itself. Nevertheless, despite our recommendation, dealing with online recruitment methods requires engaging patients fast in the recruitment process and making sure that subjects who are transferred from the digital platform are contacted and scheduled instantly for a screening visit, as the online recruitment method may be inefficient if this is not happening [69].

Unanswered Questions and Future Research

For future trials, researchers need to be able to adapt from an offline setting to an online setting, including online remedies in clinical trials, as hybrid and fully virtual trials are already emerging [70-72]. More research on online and offline recruitment strategies and what methods within the two recruitment strategies are most effective is therefore needed.

Conflicts of Interest

LEO Innovation Lab, an independently established unit of LEO Pharma A/S, employs some of the authors of this paper. This does not alter our adherence to the journal's policies.

Multimedia Appendix 1
Individual characteristics of all included studies.
[PDF File (Adobe PDF File), 1627 KB - jmir_v22i11e22179_app1.pdf]

Multimedia Appendix 2
Individual characteristics of studies included in the meta-analyses.
[PDF File (Adobe PDF File), 53 KB - jmir_v22i11e22179_app2.pdf]

Multimedia Appendix 3
Cost per enrollee for online and offline recruitment.

XSL·FO
RenderX

[PDF File (Adobe PDF File), 47 KB - jmir_v22i11e22179_app3.pdf]

References

1. Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. PLoS Med 2010 Nov 09;7(11):e1000368 [FREE Full text] [doi: 10.1371/journal.pmed.1000368] [Medline: 21085696]

2. Weiss SM. Proceedings of the National Heart and Lung Institute Working Conference on Health Behavior: Basye, Virginia, May 12-15, 1975. Bethesda, MD: National Institutes of Health; 1975.

3. Johnson O. An evidence-based approach to conducting clinical trial feasibility assessments. Clinical Investigation 2015 May;5(5):491-499 [FREE Full text] [doi: 10.4155/CLI.14.139]

4. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010 Jan;19(1-2):227-233. [doi: 10.1111/j.1365-2702.2009.03041.x] [Medline: 20500260]

5. Akers L, Gordon JS. Using Facebook for Large-Scale Online Randomized Clinical Trial Recruitment: Effective Advertising Strategies. J Med Internet Res 2018 Nov 08;20(11):e290 [FREE Full text] [doi: 10.2196/jmir.9372] [Medline: 30409765]

6. Shere M, Zhao XY, Koren G. The role of social media in recruiting for clinical trials in pregnancy. PLoS One 2014;9(3):e92744 [FREE Full text] [doi: 10.1371/journal.pone.0092744] [Medline: 24671210]

7. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]

8. Jones R, Lacroix LJ, Porcher E. Facebook Advertising to Recruit Young, Urban Women into an HIV Prevention Clinical Trial. AIDS Behav 2017 Nov;21(11):3141-3153 [FREE Full text] [doi: 10.1007/s10461-017-1797-3] [Medline: 28528463]

9. Carter-Harris L, Bartlett Ellis R, Warrick A, Rawl S. Beyond Traditional Newspaper Advertisement: Leveraging Facebook-Targeted Advertisement to Recruit Long-Term Smokers for Research. J Med Internet Res 2016 Jun 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]

10. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

11. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]

12. R: A language and environment for statistical computing. ScienceOpen. URL: https://www.scienceopen.com/document?vid=300a2dc0-3207-4383-818c-51eb0f49f561 [accessed 2020-05-22]

13. Buller DB, Meenan R, Severson H, Halperin A, Edwards E, Magnusson B. Comparison of 4 recruiting strategies in a smoking cessation trial. Am J Health Behav 2012 Sep;36(5):577-588 [FREE Full text] [doi: 10.5993/AJHB.36.5.1] [Medline: 22584086]

14. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]

15. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]

16. Cobb NK, Jacobs MA, Wileyto P, Valente T, Graham AL. Diffusion of an Evidence-Based Smoking Cessation Intervention Through Facebook: A Randomized Controlled Trial. Am J Public Health 2016 Jun;106(6):1130-1135. [doi: 10.2105/AJPH.2016.303106] [Medline: 27077358]

17. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Transl Behav Med 2015 Sep;5(3):254-259 [FREE Full text] [doi: 10.1007/s13142-015-0312-5] [Medline: 26327930]

18. Muñoz RF, Aguilera A, Schueller SM, Leykin Y, Pérez-Stable EJ. From online randomized controlled trials to participant preference studies: morphing the San Francisco Stop Smoking site into a worldwide smoking cessation resource. J Med Internet Res 2012 Jun 27;14(3):e64 [FREE Full text] [doi: 10.2196/jmir.1852] [Medline: 22739225]

19. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1:S35-S41. [doi: 10.1080/14622200601039014] [Medline: 17491169]

20. Brodar KE, Hall MG, Butler EN, Parada H, Stein-Seroussi A, Hanley S, et al. Recruiting Diverse Smokers: Enrollment Yields and Cost. Int J Environ Res Public Health 2016 Dec 16;13(12) [FREE Full text] [doi: 10.3390/ijerph13121251] [Medline: 27999280]

21. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]

22.   Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008 Dec 15;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]

23.   Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. J Med Internet Res 2012 Aug 27;14(4):e116 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]

24.   Byaruhanga J, Tzelepis F, Paul C, Wiggers J, Byrnes E, Lecathelinais C. Cost Per Participant Recruited From Rural and Remote Areas Into a Smoking Cessation Trial Via Online or Traditional Strategies: Observational Study. J Med Internet Res 2019 Nov 12;21(11):e14911 [FREE Full text] [doi: 10.2196/14911] [Medline: 31714253]

25.   Witzel TC, Gabriel MM, McCabe L, Weatherburn P, Gafos M, Speakman A, et al. Pilot phase of an internet-based RCT of HIVST targeting MSM and transgender people in England and Wales: advertising strategies and acceptability of the intervention. BMC Infect Dis 2019 Aug 08;19(1):699 [FREE Full text] [doi: 10.1186/s12879-019-4247-1] [Medline: 31391003]

26.   Bowen A, Williams M, Horvath K. Using the Internet to Recruit Rural MSM for HIV Risk Assessment: Sampling Issues. AIDS Behav 2004 Sep;8(3):311-319 [FREE Full text] [doi: 10.1023/B:AIBE.0000044078.43476.1f] [Medline: 15475678]

27.   Bull SS, Lloyd L, Rietmeijer C, McFarlane M. Recruitment and retention of an online sample for an HIV prevention intervention targeting men who have sex with men: the Smart Sex Quest Project. AIDS Care 2004 Nov;16(8):931-943. [doi: 10.1080/09540120412331292507] [Medline: 15511725]

28.   Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Del Rio C, et al. The comparability of men who have sex with men recruited from venue-time-space sampling and facebook: a cohort study. JMIR Res Protoc 2014 Jul 17;3(3):e37 [FREE Full text] [doi: 10.2196/resprot.3342] [Medline: 25048694]

29.   Buckingham L, Becher J, Voytek CD, Fiore D, Dunbar D, Davis-Vogel A, et al. Going social: Success in online recruitment of men who have sex with men for prevention HIV vaccine research. Vaccine 2017 Jun 14;35(27):3498-3505 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.002] [Medline: 28526330]

30.   Raymond HF, Rebchook G, Curotto A, Vaudrey J, Amsden M, Levine D, et al. Comparing internet-based and venue-based methods to sample MSM in the San Francisco Bay Area. AIDS Behav 2010 Feb;14(1):218-224. [doi: 10.1007/s10461-009-9521-6] [Medline: 19160034]

31.   Parsons JT, Vial AC, Starks TJ, Golub SA. Recruiting drug using men who have sex with men in behavioral intervention trials: a comparison of internet and field-based strategies. AIDS Behav 2013 Feb;17(2):688-699 [FREE Full text] [doi: 10.1007/s10461-012-0231-0] [Medline: 22684657]

32.   Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 01;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]

33.   Prescott TL, Phillips Ii G, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching Adolescent Gay, Bisexual, and Queer Men Online: Development and Refinement of a National Recruitment Strategy. J Med Internet Res 2016 Aug 04;18(8):e200 [FREE Full text] [doi: 10.2196/jmir.5602] [Medline: 27492781]

34.   Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]

35.   Maloni JA, Przeworski A, Damato EG. Web recruitment and internet use and preferences reported by women with postpartum depression after pregnancy complications. Arch Psychiatr Nurs 2013 Apr;27(2):90-95. [doi: 10.1016/j.apnu.2012.12.001] [Medline: 23540519]

36.   Adam LM, Manca DP, Bell RC. Can Facebook Be Used for Research? Experiences Using Facebook to Recruit Pregnant Women for a Randomized Controlled Trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]

37.   Ingersoll K, Frederick C, MacDonnell K, Ritterband L, Lord H, Jones B, et al. A Pilot RCT of an Internet Intervention to Reduce the Risk of Alcohol-Exposed Pregnancy. Alcohol Clin Exp Res 2018 Jun;42(6):1132-1144 [FREE Full text] [doi: 10.1111/acer.13635] [Medline: 29741798]

38.   MacDonnell K, Cowen E, Cunningham DJ, Ritterband L, Ingersoll K. Online recruitment of a non-help-seeking sample for an internet intervention: Lessons learned in an alcohol-exposed pregnancy risk reduction study. Internet Interv 2019 Sep;17:100240 [FREE Full text] [doi: 10.1016/j.invent.2019.100240] [Medline: 30963031]

39.   Guthrie KA, Caan B, Diem S, Ensrud KE, Greaves SR, Larson JC, et al. Facebook advertising for recruitment of midlife women with bothersome vaginal symptoms: A pilot study. Clin Trials 2019 Oct 06;16(5):476-480. [doi: 10.1177/1740774519846862] [Medline: 31055949]

40.   Heckman CJ, Handorf EA, Darlow SD, Ritterband LM, Manne SL. An online skin cancer risk-reduction intervention for young adults: Mechanisms of effects. Health Psychol 2017 Mar;36(3):215-225 [FREE Full text] [doi: 10.1037/hea0000420] [Medline: 27819460]

41. van Lettow B, de Vries H, Burdorf A, Boon B, van Empelen P. Drinker prototype alteration and cue reminders as strategies in a tailored web-based intervention reducing adults' alcohol consumption: randomized controlled trial. J Med Internet Res 2015 Feb 04;17(2):e35 [FREE Full text] [doi: 10.2196/jmir.3551] [Medline: 25653199]

42. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013 Feb 05;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]

43. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]

44. Jonas B, Leuschner F, Tossmann P. Efficacy of an internet-based intervention for burnout: a randomized controlled trial in the German working population. Anxiety Stress Coping 2017 Mar;30(2):133-144. [doi: 10.1080/10615806.2016.1233324] [Medline: 27602992]

45. Juraschek SP, Plante TB, Charleston J, Miller ER, Yeh H, Appel LJ, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. Clin Trials 2018 Apr;15(2):130-138 [FREE Full text] [doi: 10.1177/1740774517745829] [Medline: 29361843]

46. Bracken K, Hague W, Keech A, Conway A, Handelsman DJ, Grossmann M, et al. Recruitment of men to a multi-centre diabetes prevention trial: an evaluation of traditional and online promotional strategies. Trials 2019 Jun 19;20(1):366 [FREE Full text] [doi: 10.1186/s13063-019-3485-2] [Medline: 31217024]

47. Frandsen M, Thow M, Ferguson SG. The Effectiveness Of Social Media (Facebook) Compared With More Traditional Advertising Methods for Recruiting Eligible Participants To Health Research Studies: A Randomized, Controlled Clinical Trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]

48. Cowie JM, Gurney ME. The Use of Facebook Advertising to Recruit Healthy Elderly People for a Clinical Trial: Baseline Metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]

49. Schaub M, Sullivan R, Haug S, Stark L. Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: randomized controlled trial. J Med Internet Res 2012 Nov 28;14(6):e166 [FREE Full text] [doi: 10.2196/jmir.2244] [Medline: 23192752]

50. Schwinn T, Hopkins J, Schinke SP, Liu X. Using Facebook ads with traditional paper mailings to recruit adolescent girls for a clinical trial. Addict Behav 2017 Feb;65:207-213 [FREE Full text] [doi: 10.1016/j.addbeh.2016.10.011] [Medline: 27835860]

51. Raviotta JM, Nowalk MP, Lin CJ, Huang H, Zimmerman RK. Using Facebook™ to Recruit College-Age Men for a Human Papillomavirus Vaccine Trial. Am J Mens Health 2016 Mar 11;10(2):110-119. [doi: 10.1177/1557988314557563] [Medline: 25389213]

52. Bull SS, Vallejos D, Levine D, Ortiz C. Improving recruitment and retention for an online randomized controlled trial: experience from the Youthnet study. AIDS Care 2008 Sep;20(8):887-893. [doi: 10.1080/09540120701771697] [Medline: 18777217]

53. Das R, Machalek DA, Molesworth EG, Garland SM. Using Facebook to Recruit Young Australian Men Into a Cross-Sectional Human Papillomavirus Study. J Med Internet Res 2017 Nov 17;19(11):e389 [FREE Full text] [doi: 10.2196/jmir.8739] [Medline: 29150417]

54. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. Internet Interv 2019 Sep;17:100246 [FREE Full text] [doi: 10.1016/j.invent.2019.100246] [Medline: 31080751]

55. Carmi L, Zohar J. A comparison between print vs. internet methods for a clinical trial recruitment--a pan European OCD study. Eur Neuropsychopharmacol 2014 Jun;24(6):874-878. [doi: 10.1016/j.euroneuro.2014.01.021] [Medline: 24630432]

56. Bailey JV, Pavlou M, Copas A, McCarthy O, Carswell K, Rait G, et al. The Sexunzipped trial: optimizing the design of online randomized controlled trials. J Med Internet Res 2013 Dec 11;15(12):e278 [FREE Full text] [doi: 10.2196/jmir.2668] [Medline: 24334216]

57. Williams MT, Proetto D, Casiano D, Franklin ME. Recruitment of a hidden population: African Americans with obsessive-compulsive disorder. Contemp Clin Trials 2012 Jan;33(1):67-75 [FREE Full text] [doi: 10.1016/j.cct.2011.09.001] [Medline: 21983626]

58. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A Web-Based Cognitive Behavior Therapy Intervention to Improve Social and Occupational Functioning in Adults With Type 2 Diabetes (The SpringboarD Trial): Randomized Controlled Trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]

59. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015 Jun 01;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]

60. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb 12;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]

61. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]

62. Partridge SR, Balestracci K, Wong AT, Hebden L, McGeechan K, Denney-Wilson E, et al. Effective Strategies to Recruit Young Adults Into the TXT2BFiT mHealth Randomized Controlled Trial for Weight Gain Prevention. JMIR Res Protoc 2015 Jun 05;4(2):e66 [FREE Full text] [doi: 10.2196/resprot.4268] [Medline: 26048581]

63. Moreno MA, Waite A, Pumper M, Colburn T, Holm M, Mendoza J. Recruiting Adolescent Research Participants: In-Person Compared to Social Media Approaches. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):64-67. [doi: 10.1089/cyber.2016.0319] [Medline: 27976951]

64. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]

65. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]

66. Choi I, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: Engagement and selection bias. Internet Interv 2017 Jun;8:27-34 [FREE Full text] [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]

67. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]

68. Partlett C, Riley RD. Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. Stat Med 2017 Jan 30;36(2):301-317 [FREE Full text] [doi: 10.1002/sim.7140] [Medline: 27714841]

69. Katz B, Eiken A, Misev V, Zibert J. Optimize clinical trial recruitment with digital platforms. Dermatology Times. 2019. URL: https://www.dermatologytimes.com/view/optimize-clinical-trial-recruitment-digital-platforms [accessed 2020-02-09]

70. Clinical Trial in Subjects With Mild to Moderate Acne Vulgaris. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT02832063 [accessed 2018-06-27]

71. A Study to Evaluate the Efficacy and Safety of Rituximab Versus Mycophenolate Mofetil (MMF) in Participants With Pemphigus Vulgaris (PV). ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT02383589 [accessed 2019-10-25]

72. Clinical Trial of Patidegib Gel 2%, 4%, and Vehicle Applied Once or Twice Daily to Decrease the GLI1 Biomarker in Sporadic Nodular Basal Cell Carcinomas (BCC). ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT02828111 [accessed 2019-10-30]

Abbreviations

IR: incidence rate
IRR: incidence rate ratio
RCT: randomized controlled trial
RR: risk ratio

XSL•FO
RenderX

Original Paper

Evaluating Identity Disclosure Risk in Fully Synthetic Health Data: Model Development and Validation

Khaled El Emam[1,2,3], BEng, PhD; Lucy Mosquera[3], BSc, MSc; Jason Bass[3], BSc

[1]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[2]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
[3]Replica Analytics Ltd, Ottawa, ON, Canada

Corresponding Author:
Khaled El Emam, BEng, PhD
School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa
401 Smyth Road
Ottawa, ON, K1H 8L1
Canada
Phone: 1 6137975412
Email: kelemam@ehealthinformation.ca

Abstract

Background: There has been growing interest in data synthesis for enabling the sharing of data for secondary analysis; however, there is a need for a comprehensive privacy risk model for fully synthetic data: If the generative models have been overfit, then it is possible to identify individuals from synthetic data and learn something new about them.

Objective: The purpose of this study is to develop and apply a methodology for evaluating the identity disclosure risks of fully synthetic data.

Methods: A full risk model is presented, which evaluates both identity disclosure and the ability of an adversary to learn something new if there is a match between a synthetic record and a real person. We term this "meaningful identity disclosure risk." The model is applied on samples from the Washington State Hospital discharge database (2007) and the Canadian COVID-19 cases database. Both of these datasets were synthesized using a sequential decision tree process commonly used to synthesize health and social science data.

Results: The meaningful identity disclosure risk for both of these synthesized samples was below the commonly used 0.09 risk threshold (0.0198 and 0.0086, respectively), and 4 times and 5 times lower than the risk values for the original datasets, respectively.

Conclusions: We have presented a comprehensive identity disclosure risk model for fully synthetic data. The results for this synthesis method on 2 datasets demonstrate that synthesis can reduce meaningful identity disclosure risks considerably. The risk model can be applied in the future to evaluate the privacy of fully synthetic data.

Introduction

Data Access Challenges

Access to data for building and testing artificial intelligence and machine learning (AIML) models has been problematic in practice and presents a challenge for the adoption of AIML [1,2]. A recent analysis concluded that data access issues are ranked in the top 3 challenges faced by organizations when implementing AI [3].

A key obstacle to data access has been analyst concerns about privacy and meeting growing privacy obligations. For example, a recent survey by O'Reilly [4] highlighted the privacy concerns of organizations adopting machine learning models, with more than half of those experienced with AIML checking for privacy issues. Specific to health care data, a National Academy of Medicine/Government Accountability Office report highlights privacy as presenting a data access barrier for the application of AI in health care [5].

Anonymization is one approach for addressing privacy concerns when making data available for secondary purposes such as AIML [6]. However, there have been repeated claims of successful re-identification attacks on anonymized data [7-13], eroding public and regulator trust in this approach [13-22].

Synthetic data generation is another approach for addressing privacy concerns that has been gaining interest recently [23,24]. Different generative models have been proposed, such as decision tree–based approaches [25] and deep learning methods like Variational Auto Encoders [26,27] and Generative Adversarial Networks (GANs) [28-31].

There are different types of privacy risks. One of them is identity disclosure [23,24,32], which in our context means the risk of correctly mapping a synthetic record to a real person. Current identity disclosure assessment models for synthetic data have been limited in that they were formulated under the assumption of partially synthetic data [33-39]. Partially synthetic data permit the direct matching of synthetic records with real people because there is a one-to-one mapping between real individuals and the partially synthetic records. However, that assumption cannot be made with *fully* synthetic data whereby there is no direct mapping between a synthetic record and a real individual.

Some researchers have argued that fully synthetic data does not have an identity disclosure risk [29,40-46]. However, if the synthesizer is overfit to the original data, then a synthetic record can be mapped to a real person [47]. Since there are degrees of overfitting, even a partial mapping may represent unacceptable privacy risk. Therefore, identity disclosure is still relevant for fully synthetic data.

Another type of privacy risk is attribution risk [42,47], which is defined as an adversary learning that a specific individual has a certain characteristic. In this paper, we present a comprehensive privacy model that combines identity disclosure and attribution risk for fully synthetic data, where attribution is conditional on identity disclosure. This definition of privacy risk is complementary to the notion of membership disclosure as it has been operationalized in the data synthesis literature, where similarity between real and synthetic records is assessed [28,48]. We then demonstrate the model on health data.

Background

Key definitions and requirements will be presented, followed by a model for assessing identity disclosure risk. As a general rule, we have erred on the conservative side when presented with multiple design or parameter options to ensure that patient privacy would be less likely to be compromised.

Definitions—Basic Concepts

The basic scheme that we are assuming is illustrated in Figure 1. We have a real population denoted by the set P of size N. A real sample R exists such that $R \subseteq P$, and that is the set that we wish to create a synthetic dataset S from. Without loss of generality, the real and synthetic samples are assumed to be the same size, n.

Figure 1. The relationships between the different datasets under consideration. Matching between a synthetic sample record and someone in the population goes through the real sample and can occur in 2 directions.
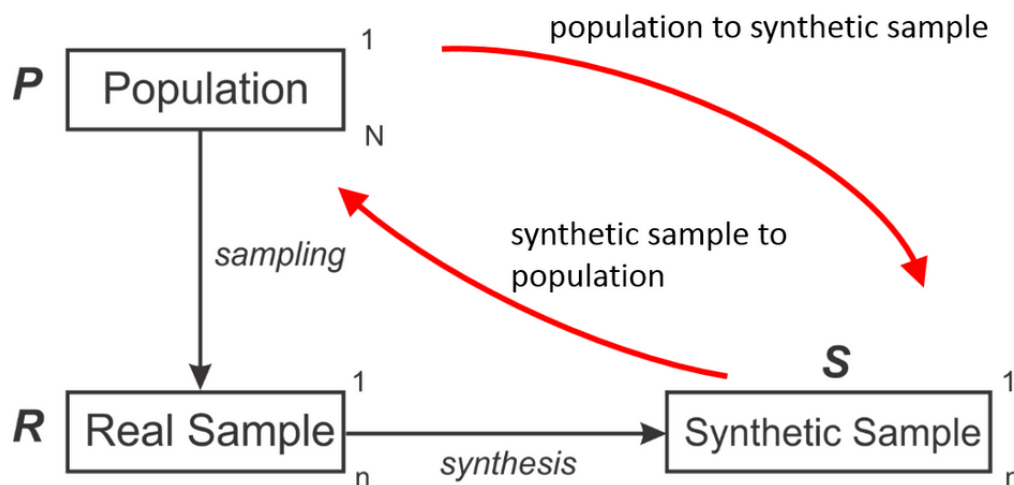
The data custodian makes the synthetic sample available for secondary purposes but does not share the generative model that is used to produce the synthetic sample. Therefore, our risk scenario is when the adversary only has access to the synthetic data.

Synthetic records can be identified by matching them with individuals in the population. When matching is performed to identify synthetic records, that matching is done on the *quasi-identifiers*, which are a subset of the variables and are known by an adversary [49]. For example, typically, a date of birth is a quasi-identifier because it is information about individuals that is known or that is relatively easy for an adversary to find out (eg, from voter registration lists [50]). More generally, an adversary may know the quasi-identifiers about an individual because that individual is an acquaintance of the adversary or because the adversary has access to a population database or registry of identifiable information.

The variables that are not quasi-identifiers will be referred to as *sensitive variables*. For example, if a dataset has information about drug use, that would be a sensitive variable that could cause harm if it was known. In general, we assume that sensitive

values would cause some degree of harm if they become known to an adversary.

To illustrate the privacy risks with fully synthetic data, consider the population data in Table 1. Individuals in the population are identifiable through their national IDs. We will treat the variable of one's origin as a quasi-identifier and one's income as the sensitive value. Table 2 displays the records from the real sample, and Table 3 presents records for the synthetic sample.

As can be seen, there is only one North African individual and one European individual in the population, and they both are in the real sample. Therefore, these unique real sample records would match 1:1 with the population and, therefore, would have a very high risk of being identified. The population-unique European and North African records are also in the synthetic data, and thus, here we have a 1:1 match between the synthetic records and the population.

The sensitive income value in the synthetic sample is very similar to the value in the real sample for the North African record. Therefore, arguably, we also learn something new about that individual. The sensitive income value is not so close for the European record, and therefore, even though we are able to match on the quasi-identifier, we will not learn meaningful information about that specific individual from synthetic data.

Table 1. Example of a population dataset, with one's origin as the quasi-identifier and one's income as the sensitive variable.

National ID	Origin	Income ($)
1	Japanese	110k
2	Japanese	100k
3	Japanese	105k
4	North African	95k
5	European	70k
6	Hispanic	100k
7	Hispanic	130k
8	Hispanic	65k

Table 2. Example of a real sample, with one's origin as the quasi-identifier and one's income as the sensitive variable.

Origin	Income ($)
European	70k
Japanese	100k
Hispanic	130k
Hispanic	65k
North African	95k

Table 3. Example of a synthetic sample, with one's origin as the quasi-identifier and one's income as the sensitive variable.

Origin	Income ($)
Japanese	115k
Japanese	120k
North African	100k
European	110k
Hispanic	65k

This example illustrates that it is plausible to match synthetic sample records with individuals in the population and thus identify these individuals, since a synthesized record can have the same value as a real record on quasi-identifiers. However, such identification is only meaningful if we learn somewhat correct sensitive information about these matched individuals. Learning something new is considered when evaluating identifiability risks in practical settings [51] and is part of the definition of identity disclosure [52]. Learning something new is also similar to the concept of attribution risk as it has been operationalized in the data synthesis literature [42,47].

Counting Matches

To formulate our model, we first need to match a synthetic sample record with a real sample record. Consider the synthetic sample in Table 3 with a single quasi-identifier, one's origin; we want to match the record with the "Hispanic" value with the real sample in Table 2. We find that there are 3 matching records in the real sample. Without any further information, we would

select one of the real sample records at random, and therefore, the probability of selecting any of the records is one-third. However, there is no correct selection here. For example, we cannot say that the third record in the real sample is the correct record match, and therefore the probability of a correct match is one-third; there is no 1:1 mapping between the fully synthetic sample records and the real sample records.

The key information here is that there was a match—it is a binary indicator. If there is a match between real sample record s and a synthetic sample record, we can use the indicator I_s (which takes on a value of 1 if there is at least one match, and 0 otherwise).

Direction of Match

A concept that is well understood in the disclosure control literature is that the probability of a successful match between someone in the population and a real record will depend on the direction of the match [53]. A randomly selected person from the real sample will always have an equivalent record in the population. However, a randomly selected record in the population may not match someone in the real sample due to sampling. The former is referred to as a sample-to-population match, and the latter as a population-to-sample match.

In our hypothetical example, an adversary may know Hans in the population and can match that with the European record in the synthetic sample through the real sample. Or the adversary may select the European record in the synthetic sample and match that with the only European in a population registry through the real sample, which happens to be Hans. Both directions of attack are plausible and will depend on whether the adversary already knows Hans as an acquaintance or not.

Now we can combine the 2 types of matching to get an overall match rate between the synthetic record and the population: the synthetic sample–to–real sample match and the real sample–to–population match, and in the other direction. We will formalize this further below.

Measuring Identification Risk

We start off by assessing the probability that a record in the real sample can be identified by matching it with an individual in the population by an adversary. The population-to-sample attack is denoted by A and the sample-to-population attack by B.

Under the assumption that an adversary will only attempt one of them, but without knowing which one, the overall probability

of one of these attacks being successful is given by the maximum of both [49]:

$$\max(A, B) \quad (1)$$

The match rate for population-to-sample attacks is given by El Emam [49] (using the notation in Table 4):

This models an adversary who selects a random individual from the population and matches them with records in the real sample. A selected individual from the population may not be in the real sample, and therefore, the sampling does have a protective effect.

Under the sample-to-population attack, the adversary randomly selects a record from the real sample and matches it to individuals in the population. The match rate is given by El Emam [49]:

We now extend this by accounting for the matches between the records in the synthetic sample and the records in the real sample. Only those records in the real sample that match with a record in the synthetic sample can then be matched with the population. We define an indicator variable, $I_s=1$, if a real sample record matches a synthetic sample record. Therefore, we effectively reduce the real sample to those records which match with at least 1 record in the synthetic sample. The population-to-synthetic sample identification risk can thus be expressed as

And similarly, the synthetic sample-to-population identification risk can be expressed as

And then we have the overall identification risk from equation (1):

The population value of $1/F$ can be estimated using methods described in various disclosure control texts [49,54-59].

Table 4. Notation used in this paper.

Notation	Interpretation
s	An index to count records in the real sample
t	An index to count records in the synthetic sample
N	The number of records in the true population
f_s	The equivalence class group size in the real sample for a particular record s in the real sample. The equivalence class is defined as the set of records with the same values on the quasi-identifiers.
F_s	The equivalence group size in the population that has the same quasi-identifier values as record s in the real sample. The equivalence class is defined as the set of records with the same values on the quasi-identifiers.
n	The number of records in the (real or synthetic) sample
I_s	A binary indicator of whether record s in the real sample matches a record in the synthetic sample
R_s	A binary indicator of whether the adversary would learn something new if record s in the real sample matches a record in the synthetic sample
k	Number of quasi-identifiers
λ	Adjustment to account for errors in matching and a verification rate that is not perfect
L	The minimal percentage of sensitive variables that need to be similar between the real sample and synthetic sample to consider that an adversary has learned something new

Adjusting for Incorrect Matches

In practice, 2 adjustments should be made to equation (6) to take into account the reality of matching when attempting to identify records [60]: data errors and the likelihood of verification. The overall probability can be expressed as:

$$pr(a)pr(b|a)pr(c|a,b)$$

$pr(a)$ is the probability that there are no errors in the data, $pr(b|a)$ is the probability of a match given that there are no errors in the data, and $pr(c|a,b)$ is the probability that the match can be verified given that there are no errors in the data and that the records match.

Real data has errors in it, and therefore, the accuracy of the matching based on adversary knowledge will be reduced [53,61]. Known data error rates not specific to health data (eg, voter registration databases, surveys, and data from data brokers) can be relatively large [62-65]. For health data, the error rates have tended to be lower [66-70], with a weighted mean of 4.26%. Therefore, the probability of at least one variable having an error in it is given by $1-(1-0.0426)^k$, where k is the number of quasi-identifiers. If we assume that the adversary has perfect information and only the data will have an error in it, then the probability of no data errors is $pr(a)=(1-0.0426)^k$.

A previous review of identification attempts found that when there is a suspected match between a record and a real individual, the suspected match could only be verified 23% of the time [71], $pr(c|a,b)=0.23$. This means that a large proportion of suspected matches turn out to be false positives when the adversary attempts to verify them. A good example from a published re-identification attack illustrating this is when the adversary was unable to contact the individuals to verify the matches in the time allotted for the study [11] (there are potentially multiple reasons for this, such as people moved, died, or their contact information was incorrect), which was 23%. It means that even though there is a suspected match,

verifying it is not certain, and without verification, it would not be known whether the match was correct. In some of these studies, the verification ability is confounded with other factors, and therefore, there is uncertainty around this 23% value.

We can now adjust equation (6) with the λ parameter:

$$\lambda = 0.23 \times (1-0.0426)^k \quad (8)$$

However, equation (8) does not account for the uncertainty in the values obtained from the literature and assumes that verification rates and error rates are independent. Specifically, when there are data errors, they would make the ability to verify less likely, which makes these 2 effects correlated. We can model this correlation, as explained below.

The verification rate and data error rate can be represented as triangular distributions, which is a common way to model phenomena for risk assessment where the real distribution is not precisely known [72]. The means of the distributions are the values noted above, and the minimum and maximum values for each of the triangular distributions were taken from the literature (cited above).

We can also model the correlation between the 2 distributions to capture the dependency between (lack of) data errors and verification. This correlation was assumed to be medium, according to Cohen guidelines for the interpretation of effect sizes [73]. We can then sample from these 2 triangular distributions inducing a medium correlation [74]. The 2 sampled values can be entered into equation (8) instead of the mean values, and we get a new value, λ_s, based on the sampled values. We draw from the correlated triangular distributions for every record in the real sample.

We can use the λ_s value directly in our model. However, to err on the conservative side and avoid this adjustment for data errors and verification over-attenuating the actual risk, we use instead the midpoint between λ_s and the maximum value of 1. We define

This more conservative adjustment can be entered into equation (6) as follows:

Learning Something New

We now extend the risk model in equation (10) to determine if the adversary would learn something new from a match. We let R_s be a binary indicator of whether the adversary could learn something new:

Because a real sample record can match multiple synthetic sample records, the R_s is equal to 1 if any of the matches meets the "learning something new" threshold.

In practice, we compute I_s first, and if that is 0, then there is no point in computing the remaining terms for that s record: we only consider those records that have a match between the real and synthetic samples since the "learning something new" test would not be applicable where there is no match.

Learning something new in the context of synthetic data can be expressed as a function of the sensitive variables. Also note that for our analysis, we assume that each sensitive variable is at the same level of granularity as in the real sample since that is the information that the adversary will have after a match.

The test of whether an adversary learns something new is defined in terms of 2 criteria: (1) Is the individual's real information different from other individuals in the real sample (ie, to what extent is that individual an outlier in the real sample)? And (2) to what extent is the synthetic sample value similar to the real sample value? Both of these conditions would be tested for every sensitive variable.

Let us suppose that the sensitive variable we are looking at is the cost of a procedure. Consider the following scenarios: If the real information about an individual is very similar to other individuals (eg, the value is the same as the mean), then the information gain from an identification would be low (note that there is still some information gain, but it would be lower than the other scenarios). However, if the information about an individual is quite different, say the cost of the procedure is 3 times higher than the mean, then the information gain could be relatively high because that value is unusual. If the synthetic sample cost is quite similar to the real sample cost, then the information gain is still higher because the adversary would learn more accurate information. However, if the synthetic sample cost is quite different from the real sample cost, then very little would be learned by the adversary, or what will be learned will be incorrect, and therefore, the correct information gain would be low.

This set of scenarios is summarized in Figure 2. Only 1 quadrant (top right) would then represent a high and correct information gain, and the objective of our analysis is to determine whether a matched individual is in that quadrant for at least L% of its sensitive variables. A reasonable value of L would need to be specified for a particular analysis.

Figure 2. The relationship between a real observation to the rest of the data in the real sample and to the synthetic observation, which can be used to determine the likelihood of meaningful identity disclosure.

		Similarity within Real Sample	
		Individual is Similar to Others	Individual is an Outlier
Similarity Between Real & Synthetic Samples	Individual's Synthetic Information Similar to Real Information	Low Meaningful Identity Disclosure Risk	High Meaningful Identity Disclosure Risk
	Individual's Synthetic Information Different from Real Information	Low Meaningful Identity Disclosure Risk	Low Meaningful Identity Disclosure Risk

This table only applies to records that match between the synthetic sample and real sample, and hence have passed the first test for what is defined as meaningful identity disclosure.

We propose a model to assess what the adversary would learn from each sensitive variable. If the adversary learns something new for at least L% of the sensitive variable, then we set $R_2=1$; otherwise, it is 0.

Nominal and Binary Sensitive Variables

We start off with nominal/binary sensitive variables and then extend the model to continuous variables. Let X_s be the sensitive variable for real record s under consideration, and let J be the set of different values that X_s can take in the real sample. Assume the matching record has value $X_s=j$ where $j \in J$, and that p_j is the proportion of records in the real sample that have the same j value.

We can then determine the distance that the X_s value has from the rest of the real sample data as follows:

$$d_j = 1 - p_j \text{ (12)}$$

The distance is low if the value j is very common, and it is large if the value of j is very different than the rest of the real sample dataset.

Let the matching record on the sensitive variable in the synthetic record be denoted by $Y_t=z$, where $z \in Z$ and Z is the set of possible values that Y_t can take in the synthetic sample; in practice, $Z \subseteq J$. For any 2 records that match from the real sample and the synthetic sample, we compare their values. The measure of how similar the real value is to the rest of the distribution when it matches is therefore given by $d_j \times [X_s = Y_t]$, where the square brackets are Iverson brackets.

How do we know if that value indicates that the adversary learns something new about the patient?

We set a conservative threshold; if the similarity is larger than 1 standard deviation, assuming that taking on value j follows a Bernoulli distribution, we then have the inequality for nominal and binary variables that must be met to declare that an adversary will learn something new from a matched sensitive variable.

The inequality compares the weighted value with the standard deviation of the proportion p_j.

Continuous Sensitive Variables

Continuous sensitive variables should be discretized using univariate k-means clustering, with optimal cluster sizes chosen by the majority rule [75]. Again, let X be the sensitive variable under consideration, and X_s be the value of that variable for the real record under consideration. We define the cluster's size in the real sample with the value of the sensitive variable that belongs to the matched real record under consideration as C_s. For example, if the sensitive variable is the cost of a procedure and it is \$150, and if that specific value is in a cluster of size 5, then $C_s=5$. The proportion of all patients that are in this cluster compared to all patients in the real sample is given by p_s.

In the same manner as for nominal and binary variables, the distance is defined as

$$d_s = p_s \quad (14)$$

Let Y_t be the synthetic value on the continuous sensitive variable that matched with real record s. The weighted absolute difference expresses how much information the adversary has learned, $d_s \times |X_s - Y_t|$.

We need to determine if this value signifies learning too much. We compare this value to the median absolute deviation (MAD) over the X variable. The MAD is a robust measure of variation. We define the inequality:

$$d_s \times |X_s - Y_t| < 1.48 \times MAD \quad (15)$$

When this inequality is met, then the weighted difference between the real and synthetic values on the sensitive variable for a particular patient indicates that the adversary will indeed learn something new.

The 1.48 value makes the MAD equivalent to 1 standard deviation for Gaussian distributions. Of course, the multiplier for MAD can be adjusted since the choice of a single standard deviation equivalent was a subjective (albeit conservative) decision.

Comprehensive Evaluation of Attacks

An adversary may not attempt to identify records on their original values but instead generalize the values in the synthetic sample and match those. The adversary may also attempt to identify records on a subset of the quasi-identifiers. Therefore, it is necessary to evaluate generalized values on the quasi-identifiers and subsets of quasi-identifiers during the matching process.

In Multimedia Appendix 1, we describe how we perform a comprehensive search for these attack modalities by considering all generalizations and all subsets, and then we take the highest risk across all combinations of generalization and quasi-identifier subsets as the overall meaningful identity disclosure risk of the dataset.

Methods

We describe the methods used to apply this meaningful identity disclosure risk assessment model on 2 datasets.

Datasets Evaluated

We apply the meaningful identity disclosure measurement methodology on 2 datasets. The first is the Washington State Inpatient Database (SID) for 2007. This is a dataset covering population hospital discharges for the year. The dataset has 206 variables and 644,902 observations. The second is the Canadian COVID-19 case dataset with 7 variables and 100,220 records gathered by Esri Canada [76].

We selected a 10% random sample from the full SID and synthesized it (64,490 patients). Then, meaningful identity disclosure of that subset was evaluated using the methodology described in this paper. The whole population dataset was used to compute the population parameters in equation (5) required for calculating the identity disclosure risk values according to equation (11). This ensured that there were no sources of estimation error that needed to be accounted for.

The COVID-19 dataset has 7 variables, with the date of reporting, health region, province, age group, gender, case status (active, recovered, deceased, and unknown), and type of exposure. A 20% sample was taken from the COVID-19 dataset (20,045 records), and the population was used to compute the meaningful identity disclosure risk similar to the Washington SID dataset.

Quasi-identifiers

State inpatient databases have been attacked in the past, and therefore, we know the quasi-identifiers that have been useful to an adversary. One attack was performed on the Washington SID [11], and a subsequent one on the Maine and Vermont datasets [10]. The quasi-identifiers that were used in these attacks and that are included in the Washington SID are shown in Table 5.

Table 5. Quasi-identifiers included in the analysis of the Washington State Inpatient Database (SID) dataset.

Variable	Definition
AGE	patient's age in years at the time of admission
AGEDAY	age in days of a patient under 1 year of age
AGEMONTH	age in months for patients under 11 years of age
PSTCO2	patient's state/county federal information processing standard (FIPS) code
ZIP	patient's zip code
FEMALE	sex of the patient
AYEAR	hospital admission year
AMONTH	admission month
AWEEKEND	admission date was on a weekend

For the COVID-19 dataset, all of the variables, except exposure, would be considered quasi-identifiers since they would be knowable about an individual.

Data Synthesis Method

For data synthesis, we used classification and regression trees [77], which have been proposed for sequential data synthesis [78] using a scheme similar to sequential imputation [79,80]. Trees are used quite extensively for the synthesis of health and social sciences data [34,81-88]. With these types of models, a variable is synthesized by using the values earlier in the sequence as predictors.

The specific method we used to generate synthetic data is called conditional trees [89], although other tree algorithms could also be used. A summary of the algorithm is provided in Textbox 1. When a fitted model is used to generate data, we sample from the predicted terminal node in the tree to get the synthetic values.

Textbox 1. Description of the sequential synthesis algorithm.

Let us say that we have 5 variables, A, B, C, D, and E. The generation is performed sequentially, and therefore, we need to have a sequence. Various criteria can be used to choose a sequence. For our example, we define the sequence as A→E→C→B→D.

Let the prime notation indicate that the variable is synthesized. For example, A' means that this is the synthesized version of A. The following are the steps for sequential generation:

- Sample from the A distribution to get A'

- Build a model F1: E ~ A

- Synthesize E as E' = F1(A')

- Build a model F2: C ~ A + E

- Synthesize C as C' = F2(A', E')

- Build a model F3: B ~ A + E + C

- Synthesize B as B' = F3(A', E', C')

- Build a model F4: D ~ A + E + C + B

- Synthesize D as D' = F4(A', E', C', B')

The process can be thought of as having 2 steps, fitting and synthesis. Initially, we are fitting a series of models (F1, F2, F3, F4). These models make up the generator. Then these models can be used to synthesize data according to the scheme illustrated above.

Risk Assessment Parameters

As well as computing the meaningful identity disclosure risk for the synthetic sample, we computed the meaningful identity disclosure risk for the real sample itself. With the latter, we let the real sample play the role of the synthetic sample, which means we are comparing the real sample against itself. This should set a baseline to compare the risk values on the synthetic data and allows us to assess the reduction in meaningful identity disclosure risk due to data synthesis. Note that both of the datasets we used in this empirical study were already de-identified to some extent.

For the computation of meaningful identity disclosure risk, we used an acceptable risk threshold value of 0.09 to be consistent with values proposed by large data custodians and have been suggested by the European Medicines Agency and Health Canada for the public release of clinical trial data (Multimedia Appendix 1). We also set L=5%.

Ethics

This study was approved by the CHEO Research Institute Research Ethics Board, protocol numbers 20/31X and 20/73X.

XSL•FO
RenderX

Results

The meaningful identity disclosure risk assessment results according to equation (11) for the Washington hospital discharge data are shown in Table 6. We can see that the overall meaningful identity disclosure risk for the synthetic data is significantly lower than the threshold of 0.09. We compare this to the real data, where the overall reduction in risk due to synthesis is approximately 5 times. The synthetic data is 4.5 times below the threshold.

The risk result on the real dataset is consistent with the empirical attack results [11]: An attempt to match 81 individuals resulted in verified, correct matches of 8 individuals, which is a risk level of 0.099 and is more or less the same as the value that was calculated using the current methodology. The real data risk was higher than the threshold, and therefore, by this standard, the original dataset would be considered to have an unacceptably high risk of identifying individuals.

The results for the synthetic Canadian COVID-19 case data are also below the threshold by about 10 times, and 4 times below risk values for the real data, although the original data has a risk value that is also below the threshold.

However, it is clear that the synthetic datasets demonstrate a significant reduction in meaningful identity disclosure risk compared to the original real dataset.

Table 6. Overall meaningful identity disclosure risk results. (The italicized values are the maximum risk values.)

Parameter	Synthetic data risk		Real data risk	
	Population-to-sample risk	Sample-to-population risk	Population-to-sample risk	Sample-to-population risk
Washington State Inpatient Database	0.00056	*0.0197*	0.016	*0.098*
Canadian COVID-19 cases	0.0043	*0.0086*	0.012	*0.034*

Discussion

Summary

The objective of this study was to develop and empirically test a methodology for the evaluation of identity disclosure risks for fully synthetic health data. This methodology builds on previous work on attribution risk for synthetic data to provide a comprehensive risk evaluation. It was then applied to a synthetic version of the Washington hospital discharge database and the Canadian COVID-19 cases dataset.

We found that the meaningful identity disclosure risk was below the commonly used risk threshold of 0.09 between 4.5 times and 10 times. Note that this reduced risk level was achieved without implementing any security and privacy controls on the dataset, suggesting that the synthetic variant can be shared with limited controls in place. The synthetic data also had a lower risk than the original data by between 4 and 5 times.

These results are encouraging in that they provide strong empirical evidence to claims in the literature that the identity disclosure risks from fully synthetic data are low. Further tests and case studies are needed to add more weight to these findings and determine if they are generalizable to other types of datasets.

Contributions of this Research

This work extends, in important ways, previous privacy models for fully synthetic data. Let R'_s be an arbitrary indicator of whether an adversary learns something new about a real sample record s. An earlier privacy risk model [42,47] focused on attribution risk was defined as:

This is similar to our definition of learning something new conditional on identity disclosure. Our model extends this work by also considering the likelihood of matching the real sample record to the population using both directions of attack, including a comprehensive search for possible matches between the real sample and synthetic sample. We also consider data errors and verification probabilities in our model, and our implementation of R'_s allows for uncertainty in the matching beyond equality tests.

Some previous data synthesis studies examined another type of disclosure: membership disclosure [28,48]. The assessment of meaningful identity disclosure, as described in this paper, does not preclude the evaluation of membership disclosure when generating synthetic data, and in fact, both approaches can be considered as complementary ways to examine privacy risks in synthetic data.

Privacy risk measures that assume that an adversary has white-box or black-box access to the generative model [29] are not applicable to our scenario, as our assumption has been that only the synthetic data is shared and the original data custodian retains the generative model.

Applications in Practice

Meaningful identity disclosure evaluations should be performed on a regular basis on synthetic data to ensure that the generative models do not overfit. This can complement membership disclosure assessments, providing 2 ways of performing a broad evaluation of privacy risks in synthetic data.

With our model, it is also possible to include meaningful identity disclosure risk as part of the loss function in generative models to simultaneously optimize on identity disclosure risk as well as data utility, and to manage overfitting during synthesis since a signal of overfitting would be a high meaningful identity disclosure risk.

Limitations

The overall risk assessment model is agnostic to the synthesis approach that is used; however, our empirical results are limited

to using a sequential decision tree method for data synthesis. While this is a commonly used approach for health and social science data, different approaches may yield different risk values when evaluated using the methodology described here.

We also made the worst-case assumption that the adversary knowledge is perfect and is not subject to data errors. This is a conservative assumption but was made because we do not have data or evidence on adversary background knowledge errors.

Future work should extend this model to longitudinal datasets, as the current risk model is limited to cross-sectional data.

Acknowledgments

Conflicts of Interest

This work was performed in collaboration with Replica Analytics Ltd. This company is a spin-off from the Children's Hospital of Eastern Ontario Research Institute. KEE is co-founder and has equity in this company. LM and JB are data scientists / software engineers employed by Replica Analytics Ltd.

Multimedia Appendix 1
Details of calculating and interpreting identity disclosure risk values.
[PDF File (Adobe PDF File), 818 KB - jmir_v22i11e23139_app1.pdf]

References

1. Government Accountability Office. Artificial Intelligence: Emerging opportunities, challenges, and implications for policy and research. U.S. GAO. 2018 Jun. URL: https://www.gao.gov/assets/700/692793.pdf [accessed 2019-07-09]
2. McKinsey Global Institute. Artificial Intelligence: The next digital frontier? McKinsey Analytics. 2017 Jun. URL: https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx [accessed 2019-07-09]
3. Deloitte. State of AI in the Enterprise, 2nd Edition. Deloitte Insights. 2018. URL: https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf [accessed 2019-07-11]
4. Lorica B, Nathan P. The State of Machine Learning Adoption in the Enterprise. Sebastopol, CA: O'Reilly; 2018.
5. Government Accountability Office, National Academy of Medicine. Artificial Intelligence in Health Care: Benefits and Challenges of Machine Learning in Drug Development (Technology Assessment). U.S. GAO. 2019 Dec. URL: https://www.gao.gov/assets/710/703558.pdf [accessed 2020-01-29]
6. Information Commissoner's Office. Anonymisation: Managing Data Protection Risk Code of Practice. ICO. 2012. URL: https://ico.org.uk/media/1061/anonymisation-code.pdf [accessed 2020-01-20]
7. de Montjoye Y, Hidalgo CA, Verleysen M, Blondel VD. Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 2013 Mar;3:1376 [FREE Full text] [doi: 10.1038/srep01376] [Medline: 23524645]
8. de MY, Radaelli L, Singh VK, Pentland AS. Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata. Science 2015 Jan 30;347(6221):536-539. [doi: 10.1126/science.1256297] [Medline: 25635097]
9. Sweeney L, Yoo JS, Perovich L, Boronow KE, Brown P, Brody JG. Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study. Technol Sci 2017:2017082801 [FREE Full text] [Medline: 30687852]
10. Su Yoo J, Thaler A, Sweeney L, Zang J. Risks to Patient Privacy: A Re-identification of Patients in Maine and Vermont Statewide Hospital Data. Technology Science 2018 Oct 08:2018100901 [FREE Full text]
11. Sweeney L. Matching Known Patients to Health Records in Washington State Data. SSRN Journal 2015 Jul 05:1-13. [doi: 10.2139/ssrn.2289850]
12. Sweeney L, von Loewenfeldt M, Perry M. Saying it's Anonymous Doesn't Make It So: Re-identifications of "anonymized" law school data. Technology Science 2018 Nov 12:2018111301 [FREE Full text]
13. 13 A. Imperiled information: Students find website data leaks pose greater risks than most people realize. Harvard John A. Paulson School of Engineering and Applied Sciences. 2020 Jan 17. URL: https://www.seas.harvard.edu/news/2020/01/imperiled-information [accessed 2020-03-23]
14. Bode K. Researchers Find "Anonymized" Data Is Even Less Anonymous Than We Thought. Motherboard: Tech by Vice. 2020 Feb 03. URL: https://www.vice.com/en_ca/article/dygy8k/researchers-find-anonymized-data-is-even-less-anonymous-than-we-thought [accessed 2020-05-11]
15. Clemons E. Online Profiling and Invasion of Privacy: The Myth of Anonymization. HuffPost. 2013 Feb 20. URL: https://www.huffpost.com/entry/internet-targeted-ads_b_2712586 [accessed 2020-05-11]

16. Jee C. You're very easy to track down, even when your data has been anonymized. MIT Technology Review. 2019 Jul 23. URL: https://www.technologyreview.com/2019/07/23/134090/youre-very-easy-to-track-down-even-when-your-data-has-been-anonymized/ [accessed 2020-05-11]

17. Kolata G. Your Data Were "Anonymized"? These Scientists Can Still Identify You. The New York Times. 2019 Jul 23. URL: https://www.nytimes.com/2019/07/23/health/data-privacy-protection.html [accessed 2020-05-05]

18. Lomas N. Researchers spotlight the lie of "anonymous" data. TechCrunch. 2019 Jul 24. URL: https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/ [accessed 2020-05-11]

19. Mitchell S. Study finds HIPAA protected data still at risks. Harvard Gazette. 2019 Mar 08. URL: https://news.harvard.edu/gazette/story/newsplus/study-finds-hipaa-protected-data-still-at-risks/ [accessed 2020-05-11]

20. Thompson S, Warzel C. Twelve Million Phones, One Dataset, Zero Privacy. The New York Times. 2019 Dec 19. URL: https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html [accessed 2020-05-11]

21. Hern A. 'Anonymised' data can never be totally anonymous, says study. The Guardian. 2019 Jul 23. URL: https://www.theguardian.com/technology/2019/jul/23/anonymised-data-never-be-anonymous-enough-study-finds [accessed 2020-05-05]

22. Ghafur S, Van Dael J, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. The Lancet Digital Health 2020 Sep;2(9):e444-e446. [doi: 10.1016/s2589-7500(20)30161-8]

23. El Emam K, Hoptroff R. The Synthetic Data Paradigm for Using and Sharing Data. Cutter Executive Update. 2019 May 06. URL: https://www.cutter.com/article/synthetic-data-paradigm-using-and-sharing-data-503526 [accessed 2020-05-06]

24. El Emam K, Mosquera L, Hoptroff R. Practical Synthetic Data Generation: Balancing Privacy and the Broad Availability of Data. Sebastopol, CA: O'Reilly Media, Inc; May 2020.

25. Reiter J. Using CART to generate partially synthetic, public use microdata. Journal of Official Statistics 2005;21(3):441-462 [FREE Full text]

26. Wan Z, Zhang Y, He H. Variational autoencoder based synthetic data generation for imbalanced learning. 2017 Presented at: IEEE Symposium Series on Computational Intelligence (SSCI); November 27 - December 1; Honolulu, Hawaii. [doi: 10.1109/ssci.2017.8285168]

27. Gootjes-Dreesbach, L, Sood M, Sahay A, Hofmann-Apitius M. Variational Autoencoder Modular Bayesian Networks (VAMBN) for Simulation of Heterogeneous Clinical Study Data. bioRxiv. 2019. URL: https://www.biorxiv.org/content/biorxiv/early/2019/09/08/760744.full.pdf [accessed 2020-01-06]

28. Zhang Z, Yan C, Mesa, DA, Sun J, Malin, BA. Ensuring electronic medical record simulation through better training, modeling, and evaluation. J Am Med Inform Assoc 2019:99-108. [doi: 10.1093/jamia/ocz161]

29. Park N, Mohammadi M, Gorde K, Jajodia S, Park H, Kim Y. Data synthesis based on generative adversarial networks. Proc. VLDB Endow 2018 Jun 01;11(10):1071-1083. [doi: 10.14778/3231751.3231757]

30. Chin-Cheong K, Sutter T, Vogt JE. Generation of Heterogeneous Synthetic Electronic Health Records using GANs. 2019 Presented at: Workshop on Machine Learning for Health (ML4H) at the 33rd Conference on Neural Information Processing Systems (NeurIPS); December 13; Vancouver. [doi: 10.3929/ethz-b-000392473]

31. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating Multi-label Discrete Patient Records using Generative Adversarial Networks. arXiv. 2017. URL: http://arxiv.org/abs/1703.06490 [accessed 2020-05-11]

32. El Emam K, Alvarez C. A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques. International Data Privacy Law 2014 Dec 13;5(1):73-87. [doi: 10.1093/idpl/ipu033]

33. Drechsler J, Reiter JP. Accounting for Intruder Uncertainty Due to Sampling When Estimating Identification Disclosure Risks in Partially Synthetic Data. In: Privacy in Statistical Databases. Lecture Notes in Computer Science, vol 5262. Berlin: Springer; 2008:227-238.

34. Drechsler J, Reiter JP. An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets. 2011 Dec;55(12):3232-3243. [doi: 10.1016/j.csda.2011.06.006]

35. Reiter JP, Mitra R. Estimating Risks of Identification Disclosure in Partially Synthetic Data. JPC 2009 Apr 01;1(1):1-1. [doi: 10.29012/jpc.v1i1.567]

36. Dandekar A, Zen R, Bressan S. A comparative study of synthetic dataset generation techniques (TRA6/18). National University of Singapore, School of Computing. 2018. URL: https://dl.comp.nus.edu.sg/bitstream/handle/1900.100/7050/TRA6-18.pdf?sequence=1&isAllowed=y [accessed 2020-07-09]

37. Loong B, Zaslavsky AM, He Y, Harrington DP. Disclosure control using partially synthetic data for large-scale health surveys, with applications to CanCORS. Statist. Med 2013 May 13;32(24):4139-4161. [doi: 10.1002/sim.5841]

38. Drechsler J, Reiter JP. Disclosure Risk and Data Utility for Partially Synthetic Data: An Empirical Study Using the German IAB Establishment Survey. Journal of Official Statistics 2008;25(4):589-603 [FREE Full text]

39. Drechsler J, Bender S, Rässler S. Comparing Fully and Partially Synthetic Datasets for Statistical Disclosure Control in the German IAB Establishment Panel. Trans. Data Privacy 2008;1(3):105-130. [doi: 10.1007/978-1-4614-0326-5_7]

40. Reiter JP. New Approaches to Data Dissemination: A Glimpse into the Future (?). CHANCE 2012 Sep 20;17(3):11-15. [doi: 10.1080/09332480.2004.10554907]

41. Hu J. Bayesian Estimation of Attribute and Identification Disclosure Risks in Synthetic Data. arXiv. 2018. URL: http://arxiv.org/abs/1804.02784 [accessed 2020-03-15]

XSL•FO
RenderX

42. Taub J, Elliot M, Pampaka M, Smith D. Differential Correct Attribution Probability for Synthetic Data: An Exploration. In: Privacy in Statistical Databases. Lecture Notes in Computer Science, vol 11126. Cham: Springer; 2018:122-137.

43. Hu J, Reiter JP, Wang Q. Disclosure Risk Evaluation for Fully Synthetic Categorical Data. In: Domingo-Ferrer J, editor. Privacy in Statistical Databases. Lecture Notes in Computer Science, vol 8744. Cham: Springer; 2014:185-199.

44. Wei L, Reiter JP. Releasing synthetic magnitude microdata constrained to fixed marginal totals. SJI 2016 Feb 27;32(1):93-108. [doi: 10.3233/sji-160959]

45. Ruiz N, Muralidhar K, Domingo-Ferrer J. On the Privacy Guarantees of Synthetic Data: A Reassessment from the Maximum-Knowledge Attacker Perspective. In: Domingo-Ferrer J, Montes F, editors. Privacy in Statistical Databases. Lecture Notes in Computer Science, vol 11126. Cham: Springer; 2018:59-74.

46. Reiter JP. Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study. J Royal Statistical Soc A 2005 Jan;168(1):185-205. [doi: 10.1111/j.1467-985x.2004.00343.x]

47. Elliot M. Final Report on the Disclosure Risk Associated with the Synthetic Data produced by the SYLLS Team. Manchester University. 2014 Oct. URL: https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/reports/2015-02%20-Report%20on%20disclosure%20risk%20analysis%20of%20synthpop%20synthetic%20versions%20of%20LCF_%20final.pdf [accessed 2020-05-11]

48. Yan C, Zhang Z, Nyemba S, Malin B. Generating Electronic Health Records with Multiple Data Types and Constraints. arXiv. 2020 Mar. URL: http://arxiv.org/abs/2003.07904 [accessed 2020-06-15]

49. El Emam K. Guide to the De-Identification of Personal Health Information. Boca Raton: CRC Press (Auerbach); 2013.

50. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]

51. Wilkinson K, Green C, Nowicki D, Von Schindler C. Less than five is less than ideal: replacing the "less than 5 cell size" rule with a risk-based data disclosure protocol in a public health setting. Can J Public Health 2020 Oct;111(5):761-765. [doi: 10.17269/s41997-020-00303-8] [Medline: 32162281]

52. Skinner C. On identification disclosure and prediction disclosure for microdata. Statistica Neerland 1992 Mar;46(1):21-32. [doi: 10.1111/j.1467-9574.1992.tb01324.x]

53. Elliot M, Dale A. Scenarios of Attack: The Data Intruders Perspective on Statistical Disclosure Risk. Netherlands Official Statistics 1999;14:6-10.

54. Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Schulte Nordholt E, Spicer K, et al. Statistical Disclosure Control. Chichester: Wiley; 2012.

55. Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Lenz R, Naylor J, et al. Handbook on Statistical Disclosure Control. ESSNet. 2010. URL: https://ec.europa.eu/eurostat/cros/system/files/SDC_Handbook.pdf [accessed 2020-01-11]

56. Duncan G, Elliot M, Salazar G. tatistical Confidentiality - Principles and Practice. New York: Springer-Verlag; 2011.

57. Templ M. Statistical Disclosure Control for Microdata. Cham: Springer International Publishing; 2017.

58. Willenborg L, de Waal T. Statistical Disclosure Control in Practice. New York: Springer-Verlag; 1996.

59. Willenborg L, de Waal T. Elements of Statistical Disclosure Control. New York: Springer-Verlag; 2001.

60. Marsh C, Skinner C, Arber S, Penhale B, Openshaw S, Hobcraft J, et al. The Case for Samples of Anonymized Records from the 1991 Census. Journal of the Royal Statistical Society. Series A (Statistics in Society) 1991;154(2):305. [doi: 10.2307/2983043]

61. Blien U, Wirth H, Muller M. Disclosure risk for microdata stemming from official statistics. Statistica Neerland 1992 Mar;46(1):69-82. [doi: 10.1111/j.1467-9574.1992.tb01327.x]

62. Inaccurate, Costly, and Inefficient: Evidence That America's Voter Registration System Needs an Upgrade. The Pew Charitable Trusts. 2012. URL: http://pew.org/2yHGTSf [accessed 2020-12-15]

63. Rainie, L, Kiesler S, Kang R, Madden M. Anonymity, Privacy, and Security Online. 2013. URL: https://www.pewresearch.org/internet/2013/09/05/anonymity-privacy-and-security-online/ [accessed 2019-12-03]

64. Leetaru, K. The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong. Forbes. 2018 Apr 05. URL: https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong/ [accessed 2019-12-03]

65. Venkatadri G, Sapiezynski P, Redmiles E, Mislove A, Goga O, Mazurek M, et al. Auditing Offline Data Brokersvia Facebook's Advertising Platform. 2019 Presented at: The World Wide Web Conference; May 13-17; San Francisco. [doi: 10.1145/3308558.3313666]

66. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. AMIA Annu Symp Proc 2008 Nov 06:242-246 [FREE Full text] [Medline: 18998889]

67. Hong MKH, Yao HHI, Pedersen JS, Peters JS, Costello AJ, Murphy DG, et al. Error rates in a clinical data repository: lessons from the transition to electronic data transfer—a descriptive study. BMJ Open 2013 May 17;3(5):e002406. [doi: 10.1136/bmjopen-2012-002406]

68. Mitchel JT, Kim YJ, Choi J, Park G, Cappi S, Horn D, et al. Evaluation of Data Entry Errors and Data Changes to an Electronic Data Capture Clinical Trial Database. Drug Information Journal 2011 Jul;45(4):421-430. [doi: 10.1177/009286151104500404]

69. Wahi MM, Parks DV, Skeate RC, Goldin SB. Reducing Errors from the Electronic Transcription of Data Collected on Paper Forms: A Research Data Case Study. Journal of the American Medical Informatics Association 2008 May 01;15(3):386-389. [doi: 10.1197/jamia.m2381]

70. Nahm ML, Pieper CF, Cunningham MM. Quantifying Data Quality for Clinical Trials Using Electronic Data Capture. PLoS ONE 2008 Aug 25;3(8):e3049. [doi: 10.1371/journal.pone.0003049]

71. Branson J, Good N, Chen J, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. Trials 2020 Feb 18;21(1):1-1. [doi: 10.1186/s13063-020-4120-y]

72. Vose D. Risk Analysis: A Quantitative Guide, 3rd ed. Chichester: Wiley; 2008.

73. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J: Lawrence Erlbaum Associates; 1988.

74. Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. Communications in Statistics - Simulation and Computation 2007 Jun 27;11(3):311-334. [doi: 10.1080/03610918208812265]

75. Charrad M, Ghazzali N, Boiteau V, Niknafs A. : An Package for Determining the Relevant Number of Clusters in a Data Set. J. Stat. Soft 2014;61(6):1-1. [doi: 10.18637/jss.v061.i06]

76. Esri Canada. Covid-19 Resources. Covid-19 Canada. URL: https://resources-covid19canada.hub.arcgis.com/ [accessed 2020-10-15]

77. Gordon AD, Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Biometrics 1984 Sep;40(3):874. [doi: 10.2307/2530946]

78. Reiter J. Using CART to generate partially synthetic, public use microdata. Journal of Official Statistics 2005;21(3):441-462.

79. Conversano C, Siciliano R. Incremental Tree-Based Missing Data Imputation with Lexicographic Ordering. J Classif 2010 Jan 7;26(3):361-379. [doi: 10.1007/s00357-009-9038-8]

80. Conversano C, Siciliano R. Tree based classifiers for conditional incremental missing data imputation. Department of Mathematics and Statistics, University of Naples. 2002. URL: http://erin.it.jyu.fi/dataclean/abstracts/node25.html [accessed 2020-05-11]

81. Arslan RC, Schilling KM, Gerlach TM, Penke L. Using 26,000 diary entries to show ovulatory changes in sexual desire and behavior. J Pers Soc Psychol 2018 Aug 27:1-48. [doi: 10.1037/pspp0000208] [Medline: 30148371]

82. Bonnéry D, Feng Y, Henneberger AK, Johnson TL, Lachowicz M, Rose BA, et al. The Promise and Limitations of Synthetic Data as a Strategy to Expand Access to State-Level Multi-Agency Longitudinal Data. Journal of Research on Educational Effectiveness 2019 Aug 02;12(4):616-647. [doi: 10.1080/19345747.2019.1631421]

83. Sabay A, Harris L, Bejugama V, Jaceldo-Siegl K. Overcoming Small Data Limitations in Heart Disease Prediction by Using Surrogate Data. SMU Data Science Review 2018;1(3):12 [FREE Full text]

84. Freiman M, Lauger A, Reiter J. Data Synthesis and Perturbation for the American Community Survey at the U.S. Census Bureau. US Census Bureau. 2017. URL: https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/2017%20Data%20Synthesis%20and%20Perturbation%20for%20ACS.pdf [accessed 2020-05-05]

85. Nowok B. Utility of synthetic microdata generated using tree-based methods. Administrative Data Research Centre, University of Edinburgh. 2015. URL: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/20150/Paper_33_Session_2_-_Univ._Edinburgh__Nowok_.pdf [accessed 2020-05-11]

86. Raab GM, Nowok B, Dibben C. Practical Data Synthesis for Large Samples. JPC 2018 Feb 02;7(3):67-97. [doi: 10.29012/jpc.v7i3.407]

87. Nowok B, Raab GM, Dibben C. Providing bespoke synthetic data for the UK Longitudinal Studies and other sensitive data with the synthpop package for R1. SJI 2017 Aug 21;33(3):785-796. [doi: 10.3233/sji-150153]

88. Quintana DS. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. eLife 2020 Mar 11;9:e53275 [FREE Full text] [doi: 10.7554/eLife.53275] [Medline: 32159513]

89. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics 2006 Sep;15(3):651-674. [doi: 10.1198/106186006X133933]

Abbreviations

AIML: artificial intelligence and machine learning
MAD: median absolute deviation
SID: State Inpatient Database

XSL•FO

RenderX

Proposal and Assessment of a De-Identification Strategy to Enhance Anonymity of the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) in a Public Cloud-Computing Environment: Anonymization of Medical Data Using Privacy Models

Seungho Jeon[1], ME; Jeongeun Seo[1], BD; Sukyoung Kim[1], BD; Jeongmoon Lee[2], BD; Jong-Ho Kim[3], PhD; Jang Wook Sohn[4], MD, PhD; Jongsub Moon[1*], PhD; Hyung Joon Joo[5*], MD, PhD

[1]Division of Information Security, Graduate School of Information Security, Korea University, Seoul, Republic of Korea

[2]Korea University Research Institute for Medical Bigdata Science, Korea University, Seoul, Republic of Korea

[3]Department of Cardiology, Cardiovascular Center, Korea University, Seoul, Republic of Korea

[4]Division of Infectious Diseases, Department of Internal Medicine, College of Medicine, Korea University, Seoul, Republic of Korea

[5]Department of Internal Medicine, Korea University College of Medicine, Korea University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyung Joon Joo, MD, PhD
Department of Internal Medicine
Korea University College of Medicine
Korea University
145 Anam-ro, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 10 3476 0525
Email: drjoohj@gmail.com

Abstract

Background: De-identifying personal information is critical when using personal health data for secondary research. The Observational Medical Outcomes Partnership Common Data Model (CDM), defined by the nonprofit organization Observational Health Data Sciences and Informatics, has been gaining attention for its use in the analysis of patient-level clinical data obtained from various medical institutions. When analyzing such data in a public environment such as a cloud-computing system, an appropriate de-identification strategy is required to protect patient privacy.

Objective: This study proposes and evaluates a de-identification strategy that is comprised of several rules along with privacy models such as k-anonymity, l-diversity, and t-closeness. The proposed strategy was evaluated using the actual CDM database.

Methods: The CDM database used in this study was constructed by the Anam Hospital of Korea University. Analysis and evaluation were performed using the ARX anonymizing framework in combination with the k-anonymity, l-diversity, and t-closeness privacy models.

Results: The CDM database, which was constructed according to the rules established by Observational Health Data Sciences and Informatics, exhibited a low risk of re-identification: The highest re-identifiable record rate (11.3%) in the dataset was exhibited by the DRUG_EXPOSURE table, with a re-identification success rate of 0.03%. However, because all tables include at least one "highest risk" value of 100%, suitable anonymizing techniques are required; moreover, the CDM database preserves the "source values" (raw data), a combination of which could increase the risk of re-identification. Therefore, this study proposes an enhanced strategy to de-identify the source values to significantly reduce not only the highest risk in the k-anonymity, l-diversity, and t-closeness privacy models but also the overall possibility of re-identification.

Conclusions: Our proposed de-identification strategy effectively enhanced the privacy of the CDM database, thereby encouraging clinical research involving multiple centers.

Introduction

The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), defined by the nonprofit organization Observational Health Data Sciences and Informatics (OHDSI) [1], is a standard data schema [2,3] that uses standardized terms [4]. An established CDM database can be used by multiple institutions to conduct quick analyses under the same conditions using the analysis tools provided by OHDSI, such as Atlas [5] and Achilles [6]. Furthermore, CDM can be used not only for research in combination with a distributed research network for collecting results analyzed by an individual institution but also in clinical decision support systems for patient-specific medical treatment by combining advanced analysis and prediction techniques such as artificial intelligence. Consequently, many medical institutions have recently attempted establishing CDM databases.

However, owing to security concerns, existing CDM databases remain inaccessible from outside the network of their institutions; therefore, only authorized researchers are allowed to analyze their clinical data. Moreover, institutions operate the CDM databases only in on-premise environments for security concerns regarding sensitive information. Nevertheless, owing to increasing system complexities and service availability, operators generally prefer to run the CDM database in a cloud-computing environment. This recent trend has led to recent regulatory and legal considerations regarding network accessibility [7].

Meanwhile, while the demand for research using medical data that has been accumulated over the past decades has increased, interinstitutional research specifically requires appropriate de-identification of medical data because of privacy concerns. Hence, to address these concerns, anonymization studies and frameworks are currently being proposed for various datasets [8-10].

As CDM database research does not extract nor analyze the institutional raw data, it involves a low risk of personal information disclosure. However, as various clinical databases are accessed and analyzed in a public environment, the construction of a CDM database requires not only an access control policy but also highly tailored protection mechanisms in addition to evaluation of the adequacy of the anonymization [11]. Therefore, in this study, we proposed and evaluated a de-identification strategy for the OMOP-CDM database using privacy models such as k-anonymity.

Methods

OMOP-CDM Database

The OMOP-CDM database is one of the significant core projects managed by the OHDSI, which uses common representation for the clinical data of various projects to support medical research. For instance, ATHENA is a standardized vocabulary [4], Atlas is a unified web-based interface to analyze CDM data [5], and Achilles is used for data characterization [6]; in addition to these tools, several useful applications have also been provided [12-14].

The specifications of the OMOP-CDM [15] are actively being amended according to the needs of researchers. At the time of writing this article, version 6.0.0 of the CDM had been published. However, this research uses the CDM database constructed by the Korea University Anam Hospital on March 23, 2020, based on version 5.3.1 of the OMOP-CDM schema. The tables comprising the database include the standardized vocabulary, metadata, clinical data tables, health system data tables, health economics, and derived elements, in addition to the results schema.

The standardized vocabulary contains 10 tables with detailed information on the concepts used in the OMOP-CDM, while the standardized metadata preserves the entire metadata information derived from data that have been transformed into the OMOP-CDM database in 2 tables. Similarly, the 16 standardized clinical tables contain patient clinical data; the tables that store the clinical data have a relation with the PERSON table, which stores the patient's personal information; the 4 standardized health system data tables contain information on the agency providing the treatment; the 2 standardized health economics data tables contain payment information for the medical services; the standardized derived elements store information such as the dosing period in 3 tables; and the results schema stores information such as the definition of each cohort in 2 tables. In the OMOP-CDM, some raw source data collected by institutions remain in fields named "source_value," which are used for flexibility and research convenience and are generally not shared outside the source institution. However, because of the high accessibility of the cloud-computing environment, these should be protected using proper security measures.

ARX Anonymization Framework

This study uses the ARX anonymization framework [16] as the evaluation tool for de-identification. ARX is awidely used open-source software for anonymizing data that include sensitive information; it supports a variety of privacy models, risk analysis models, data transformation methods, and methods to analyze the anonymized data.

The anonymizing process using ARX can be divided into 4 steps: (1) configuration, (2) exploration, (3) utility analysis, and (4) risk analysis. In the configuration step, ARX imports the data and then assigns each field to one of the following elements: identifier, quasi-identifier, sensitive data, or insensitive data. The information in the fields that are set as identifiers is removed in the anonymizing process. Transformation rules such as generalization or aggregation are applied to the data in the fields that are set as quasi-identifiers: The fields marked as sensitive

XSL•FO

RenderX

or insensitive data do not undergo transformation, and the sensitive fields are protected by a privacy model such as k-anonymity. Once the field attribution is complete, ARX creates the privacy model and sets the model parameters before finally performing the data anonymization. In addition, ARX represents possible transformations as a lattice and can produce various levels of data anonymization depending on the transformation rules and the selected privacy model. In the exploration step, the user explores the results and selects the transformation for the analysis. In the utility analysis step, the user evaluates the transformation selected in the previous step by various statistical analysis methods such as logistic regression and chooses a suitable transformation for the de-identification scenario expected by the user. Finally, in the risk analysis step, ARX analyzes the risk of re-identification under the chosen transformation using 3 re-identification attacker models: prosecutor, journalist, and marketer [17]. In the prosecutor attacker model, the attacker is assumed to know the targeted individual in the database. In the journalist attacker model, an attacker does not know whether the targeted person is listed in the database. In the marketer attacker model, an attacker aims at identifying a large number of records from a database, rather than any particular individual. This study excluded the CDM tables that did not undergo the extract-transform-load (ETL) procedure from the evaluation process and randomly sampled the data from the database for tables exceeding the maximum analytical data size of ARX (*rows* x *cols* $= 2^{32} - 1$).

Methods for Anonymizing Personally Identifiable Information

Anonymization techniques for personally identifiable information include pseudonymization, aggregation, data reduction, categorization, and masking. Pseudonymization is the generic term for replacing personally identifiable information with other values; it includes heuristic pseudonymization, encryption, and substitution. In aggregation, an individual value is replaced with a statistical value such as the mean or median of the identifiers or quasi-identifiers of the group to which the individual belongs. Data reduction is the simple removal of elements that make an individual identifiable: Although this is the strongest anonymizing method, it causes loss of some important data and degrades the value of the entire dataset. Categorization is the most common anonymization method, wherein personal information is replaced with a representative group value. Finally, masking converts a portion of the data to either a blank or noise, such as "*".

Results

Proposal for De-Identification of Personal Information in CDM

Owing to the importance of de-identifying personal information while using personal health data for secondary research, the OMOP-CDM already implements a certain level of de-identification during the construction of its database: The reference architecture provided by OHDSI, namely the OHDSIonAWS [18], uses several anonymization methods to comply with the Health Information Portability and Accountability Act (HIPAA) [7]. However, although the OMOP-CDM adopts standardized terminology, it is designed to maintain the original source information in the field of "_SOURCE_VALUE." Therefore, it is necessary to consider the appropriateness of the de-identification level for this database. In this study, we propose an enhanced de-identification strategy that is comprised of a set of rules for privacy models such as k-anonymity, l-diversity, and t-closeness for the OMOP-CDM from the perspective of reconnection with other information and privacy models.

The de-identification strategy for the CDM presented in this research is as follows: First, according to the Safe Harbor method in the HIPAA privacy rule [19], the data included in the CDM are classified as identifiers, quasi-identifiers, raw data, sensitive data, or insensitive data. "Raw data" refers to fields including the keyword "source_value" in the OMOP-CDM convention: A raw data field preserves data from the institution's database before it is converted to the CDM database through ETL. Second, the identifiers are deleted, and the raw data are protected by anonymization methods; specifically, the foreign key that can be linked with the institutional database is deleted. Third, fields containing intentionally stored data such as those in the SURVEY_CONDUCT table are maintained. Fourth, if a field includes identifiers or quasi-identifiers, this field or table must not be used, although it is intentionally stored.

Table 1 shows the transformations for the fields named "source_value." These rules collaborate with the privacy models to achieve a high de-identification level. The table lists only the fields named by the "source_value" or those possibly containing (quasi-)identifiers.

Identifiers such as the name and social security number are removed during the ETL. We designated the fields in Table 1 as quasi-identifiers because they contain raw data. However, the fields containing clinical data such as "quantity" in DRUG_EXPOSURE are considered as sensitive information and are therefore protected by anonymizing the quasi-identifiers using either the l-diversity or t-closeness model.

In Table 1, most tables such as PERSON and OBSERVATION clearly contain sensitive information that may identify a specific individual. However, the data stored in the LOCATION, CARE_SITE, PROVIDER, PAYER_PLAN, and COST fields do not directly identify the individual: For instance, the CARE_SITE table contains information regarding the institution at which health care has been delivered. Despite its property, this information may be sensitive because a combination of these data could implicitly specify the individual. However, following the proposed strategy could reduce the number of data records. For instance, even after de-identification is implemented, if the combination of the identifiers and quasi-identifiers is unique or does not satisfy the pre-defined criteria of the privacy models (ie, "k," "l," or "t"), this data could identify the individual in the database. Although there seems to be a loss of important data, this strategy actually minimizes the impact on the data analyses from de-identification because the "source_value" fields can be replaced by the "source_concept_id" fields. For instance, the "gender_source_value" of the PERSON table stores the gender of the individuals as recorded in the institution: For example,

this value could be "Male/Female" or "0/1"; the standard   concept for gender in OMOP-CDM is "M/F."

Table 1. Proposal for enhanced de-identification for the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) specification. Information regarding the CDM table, its fields, and its descriptions have been cited from [15].

Tables and fields	Description	De-identification method
PERSON		
Person_source_value	A key derived from a personal identifier in the source data	Remove[a]
Gender_source_value, Race_source_value, Ethnicity_source_value	The codes for the gender, race, and ethnicity of an individual as they appear in the source data	Masking[b]
VISIT_OCCURRENCE		
Visit_source_value, Admitting_from_source_value, Discharge_to_source_value	The codes for the visit, where the patient was admitted from, and the discharge disposition as they appear in the source data	Masking
VISIT_DETAIL		
Visit_detail_source_value, Admitted_from_source_value, Discharge_to_source_value	The codes for the visit, admitting source, discharge disposition, and optional details information as they appear in the source data	Masking
CONDITION_OCCURRENCE		
Condition_source_value, Condition_status_source_value	The codes for the condition and its status as they appear in the source data	Masking
DRUG_EXPOSURE		
Sig	The directions on the drug prescription as recorded in the original prescription	NLP[c]
Drug_source_value, Route_source_value, Dose_unit_source_value	The codes for the drug, administration route, and dose unit as they appear in the source data	Masking
PROCEDURE_OCCURRENCE		
Procedure_source_value, Modifier_source_value	The codes for the procedure and qualifier as they appear in the source data	Masking
DEVICE_EXPOSURE		
Device_source_value	The code for the device as it appears in the source data	Masking
MEASUREMENT		
Measurement_source_value, Unit_source_value, Value_source_value	The measurement name, unit (code), and value as a number as they appear in the source data	Masking
NOTE		
Note_text	The content of the note	NLP
SURVEY_CONDUCT		
Validated_survey_source_value, Survey_source_identifier	Source value representing the validation status of the survey and a unique identifier for each completed survey in the source system	Masking
OBSERVATION		
Value_as_string	The observation result stored as a string	NLP
SPECIMEN		
Specimen_source_value, Unit_source_value, Anatomic_site_source_value, Disease_status_source_value	The specimen value, unit information, anatomic site, and disease status information as they appear in the source data	Masking
LOCATION		
Address_1, address_2	The address field (street address, building, suite, floor) and zip or postal code as they appear in the source data	Partial remove[d]
Latitude, longitude	The geocoded latitude and longitude	Remove
Location_source_value	The information that is used to uniquely identify the location as it appears in the source data	Masking
CARE_SITE		
Care_site_source_value	The identifier for the care site and the source code for the Place of Service in the source data	Masking

XSL•FO
RenderX

Tables and fields	Description	De-identification method
PROVIDER		
Provider_name, npi, dea, Provider_source_value, Specialty_source_value, Gender_source_value	The provider name, National Provider Identifier (NPI), Drug Enforcement Administration (DEA) number of the provider, provider identifier, source code for the provider specialty, and gender information in the source data	Masking
PAYER_PLAN_PERIOD		
Payer_source_value, Plan_source_value, Contract_source_value, Sponsor_source_value, Family_source_value	The source codes for the payer, health benefit plan, reason justifying the contract, sponsor of the health plan, family, and reason for stopping coverage as they appear in the source data	Masking
COST		
Cost_source_value, Revenue_code_source_value, Drg_source_value	The source values for the cost, revenue code, and three-digit drug source code as they appear in the source data	Masking

[a]Refers to the deletion of the entire value stored in the field.

[b]Refers to the replacement of a part of the value with another character such as "*".

[c]NLP: natural language processing.

[d]Refers to deletion of only a part of the value.

Evaluation and Validation of De-Identification of Personal Information in CDM

The aforementioned principles apply to identifiers, quasi-identifiers, and raw data. However, personal information that is often disclosed as sensitive information must also be protected via an appropriate de-identification policy according to the laws in force like HIPAA as well as the adopted privacy models. As already mentioned, the commonly used privacy models include k-anonymity [20], l-diversity [21], and t-closeness [22], among which k-anonymity is most widely used. In this model, anonymization methods are first applied to the identifiers and quasi-identifiers. Then, all records are divided into groups such that each group includes all records with identical anonymized identifiers and quasi-identifiers; a group containing fewer than k records is discarded. Therefore, the probability of identifying an individual is $1/k$. However, if the anonymization is applied only to the identifiers and quasi-identifiers and the sensitive information is not hidden or anonymized, then an individual can be identified using the personal information that is already known — this is called a homogeneity attack. To reduce the risk of such attacks, the l-diversity model is used, which further divides the groups in the k-anonymity model such that the elements in each group have at least l different values for the sensitive data; groups with fewer than l elements are deleted. Although the application of this model can improve the de-identification level, some individual information exposure is still possible: For instance, if the distribution of sensitive data in each group is significantly skewed (ie, the sensitive data hold biased information in a particular space [or domain]), it is possible to deduce that an individual is associated with that domain — this is called a skewness attack. Therefore, to protect against such attacks, the t-closeness model is introduced, which forces domain distribution between groups of less than or equal to a pre-defined t. In this model, the distribution of the sensitive data in each group occurs between similar areas. The choice of the

appropriate privacy model can be made by considering the value of the dataset and impact in the event of data disclosure.

As all 3 privacy models essentially achieve their security goals by anonymizing the identifiers and quasi-identifiers, the personal information administrator should appropriately designate identifiers or quasi-identifiers in a dataset following the institutional security policy.

In this research, we set the identifiers and quasi-identifiers to establish the transformation rules according to the proposed strategy to improve the de-identification level of the OMOP-CDM. The following is a detailed explanation: The proposed de-identification strategy can be achieved by using one of the privacy models combined with the rules described in Table 1. If the administrator chooses the k-anonymity privacy model, the fields in Table 1 are considered as quasi-identifiers and anonymized to stand for a pre-defined group of size k. However, if the l-diversity or t-closeness model is chosen, the fields in Table 1 are regarded as quasi-identifiers like in the former case, while the other fields containing clinical data, such as "quantity" in DRUG_EXPOSURE, are designated as sensitive data. Then, the anonymizing process proceeds for quasi-identifiers while complying with pre-defined "l" or "t."

Table 2 presents the number of data values for each table included in the CDM database established by the Korea University Anam Hospital that was used for the evaluation, wherein we analyzed the re-identification risk for the individuals in the PERSON table as well as those joined to the PERSON table through the other tables including the patient-level clinical data. While anonymizing personal information using ARX, each field in the table should be designated as the identifier, quasi-identifier, sensitive data, or insensitive data. As the identifier has already been removed from the OMOP-CDM schema during the ETL process, each field in the table is classified as 1 of the 3 remaining types. In our proposed strategy, the raw data are considered as quasi-identifiers and are therefore anonymized by the transformation rules when privacy models, such as k-anonymity, l-diversity, and t-closeness, are applied.

In this evaluation, the fields listed in Table 1 were designated as quasi-identifiers, while the remaining fields, including patient-level clinical data other than the PERSON table, were designated as sensitive data.

Table 2. Number of data values per table in the Korea University Anam Hospital Common Data Model (CDM) database.

Tables in the database	Number of values
PERSON	1,891,755
CONDITION_OCCURRENCE	28,704,247
CONDITION_ERA	14,972,790
DEVICE_EXPOSURE	33,617,896
DOSE_ERA	64,047,133
DRUG_ERA	29,274,258
DRUG_EXPOSURE	77,919,053
MEASUREMENT	196,567,735
OBSERVATION	1,744,021
OBSERVATION_PERIOD	1,629,356
PROCEDURE_OCCURRENCE	21,200,346

Tables 3, 4, and 5 present the results of the re-identification risk analysis when the k-anonymity, l-diversity, and t-closeness privacy models, respectively, are applied to the CDM database. This means that the presented de-identification strategy is applied for the fields containing the source data in each table after ETL. In these tables, the re-identification risk indicates that some combinations of source data uniquely appear in the CDM database; this further implies re-identification for the individual. The ARX supports 3 re-identification attacker models for the re-identification risk analysis, namely the prosecutor, journalist, and marketer models. Herein, we used the prosecutor model, which analyzes the possibility of an attacker identifying a person under the assumption that they are already aware that the person is included in the dataset. Some fields requiring natural language processing (NLP) have been removed because

ARX does not support the anonymizing techniques for free text. Furthermore, even if free text is appropriately anonymized, it is likely to have unique values, which make it difficult to meet the criteria of the privacy models used in this study. The tables present the results of applying the optimal transformation suggested by ARX. In accordance with the applicable national guidelines for de-identification of personal information [23], k-anonymity was applied with the minimum criterion, $k = 3$, while the remaining models used the most conservative conditions, guaranteeing stronger anonymity than the k-anonymity model (ie, $l = 5$, $t = 0.1$). For instance, if l is <5 in the l-diversity model applied to this dataset, ARX suggests a level of transformation equal to that of k-anonymity for most of the tables.

Table 3. Re-identification risks before and after source_value anonymization (k-anonymity, k = 3).

Tables	Records at risk[a] (%)	Highest risk[b] (%)	Success rate[c] (%)
PERSON			
Before[d]	0.1	100	0.11
After[e]	0	1.27	0.01
CONDITION_OCCURRENCE			
Before	7.58	100	6.39
After	<0.01	25	<0.01
CONDITION_ERA			
Before	0.02	100	0.03
After	0	6.25	<0.01
DEVICE_EXPOSURE			
Before	6.98	100	5.73
After	0	0.3	<0.01
DOSE_ERA			
Before	0.02	100	0.03
After	<0.01	25	<0.01
DRUG_ERA			
Before	0.02	100	0.03
After	0	16.67	<0.01
DRUG_EXPOSURE			
Before	11.31	100	8.86
After	<0.01	33.33	<0.01
MEASUREMENT			
Before	1.53	100	1.56
After	0	6.25	<0.01
OBSERVATION			
Before	1.63	100	1.55
After	0	0.98	<0.01
OBSERVATION_PERIOD			
Before	0.11	100	0.12
After	0	6.25	<0.01
PROCEDURE_OCCURRENCE			
Before	4.25	100	3.7
After	0	0.37	<0.01

[a]Represents the percentage of data in the dataset that exceeds the risk threshold.

[b]Represents the highest risk for an individual data value.

[c]Percentage of data that can be re-identified in the dataset on average.

[d]Refers to before applying the anonymizing methods to the data.

[e]Refers to after applying the anonymizing methods to the data.

XSL•FO
RenderX

Table 4. Re-identification risks before and after source_value anonymization (l-diversity, l = 5).

Tables	Records at risk[a] (%)	Highest risk[b] (%)	Success rate[c] (%)
CONDITION_OCCURRENCE			
Before[d]	7.58	100	6.39
After[e]	0	0.36	<0.01
CONDITION_ERA			
Before	0.02	100	0.03
After	0	0.38	<0.01
DEVICE_EXPOSURE			
Before	6.98	100	5.73
After	0	0.3	<0.01
DOSE_ERA			
Before	0.02	100	0.03
After	0	0.2	<0.01
DRUG_ERA			
Before	0.02	100	0.03
After	0	0.2	<0.01
DRUG_EXPOSURE			
Before	11.31	100	<8.86
After	0	0.19	<0.01
MEASUREMENT			
Before	1.53	100	1.56
After	0	<0.01	<0.01
OBSERVATION			
Before	1.63	100	1.55
After	0	0.08	0.01
OBSERVATION_PERIOD			
Before	0.11	100	0.12
After	0	0.58	<0.01
PROCEDURE_OCCURRENCE			
Before	4.25	100	3.7
After	0	0.37	<0.01

[a]Represents the percentage of data in the dataset that exceeds the risk threshold.

[b]Represents the highest risk for an individual data value.

[c]Percentage of data that can be re-identified in the dataset on average.

[d]Refers to before applying the anonymizing methods to the data.

[e]Refers to after applying the anonymizing methods to the data.

Table 5. Re-identification risks before and after source_value anonymization (t-closeness, $t = 0.1$).

Tables	Records at risk[a] (%)	Highest risk[b] (%)	Success rate[c] (%)
CONDITION_OCCURRENCE			
Before[d]	7.58	100	6.39
After[e]	0	<0.01	<0.01
CONDITION_ERA			
Before	0.02	100	0.03
After	0	<0.01	<0.01
DEVICE_EXPOSURE			
Before	6.98	100	5.73
After	0	<0.01	<0.01
DOSE_ERA			
Before	0.02	100	0.03
After	0	<0.01	<0.01
DRUG_ERA			
Before	0.02	100	0.03
After	0	<0.01	<0.01
DRUG_EXPOSURE			
Before	11.31	100	8.86
After	0	<0.01	<0.01
MEASUREMENT			
Before	1.53	100	1.56
After	0	<0.01	<0.01
OBSERVATION			
Before	1.63	100	1.55
After	0	<0.01	<0.01
OBSERVATION_PERIOD			
Before	0.11	100	0.12
After	0	<0.01	<0.01
PROCEDURE_OCCURRENCE			
Before	4.25	100	3.7
After	0	<0.01	<0.01

[a]Represents the percentage of data in the dataset that exceeds the risk threshold.

[b]Represents the highest risk for an individual data value.

[c]Percentage of data that can be re-identified in the dataset on average.

[d]Refers to before applying the anonymizing methods to the data.

[e]Refers to after applying the anonymizing methods to the data.

As presented in the analysis results, before applying the de-identification strategy (ie, immediately after the ETL), the CDM database exhibits a significantly low possibility of re-identification. The DRUG_EXPOSURE table presented the highest percentage of records at risk, at 11.31% (8,812,644 records), and an average re-identification attack success rate of 8.86% (6,903,628 records) through a combination of the source values. The DOSE_ERA table yielded the lowest percentage of records at risk, at 0.02% (12,809 records), and an average re-identification attack success rate of 0.03% (19,214 records). Although the OMOP-CDM already had a high level of anonymity, it exhibited the highest risk (100%) for re-identifiable data in all tables. In other words, there is a risk of an individual being identified solely by a combination of the "source values." Therefore, a de-identification strategy for the "source value" fields is required for the safe utilization of the CDM data in a public environment such as a cloud-computing system. In fact, in every case where the privacy model was

applied, the "highest risks" were substantially reduced, and the overall re-identification possibility was also reduced. Among the privacy models, the k-anonymity model demonstrated the weakest de-identification level, while the t-closeness model yielded the strongest de-identification results. These results are expected because l-diversity and t-closeness models are stronger and complement the k-anonymity model, as we had set conservatively stronger configurations for these 2 models.

Impact of De-Identification on Data Analysis

We conducted several experiments to empirically observe the impact of the proposed de-identification strategy on the analysis: The analysis codes generated by Atlas were used in these experiments. We arbitrarily selected 2 CDM-based studies published on the internet [24,25] to evaluate the proposed de-identification strategy. All parameters of the 2 OMOP-CDM databases used in the experiments were set to the same values, except for the data to be de-identified.

Figure 1 shows the cohorts defined in the "Influenza Cohort Diagnostics" [24] in the "Covid19Hospitalization Characterization" study [26] before and after the proposed de-identification. This study describes the baseline characteristics of influenza (H1N1) patients between 2009-2010 and 2014-2019, builds several cohorts, and systematically presents the characteristics of patients with influenza according to age or gender. Obviously, the same cohorts were generated according to the same definitions before and after the de-identification. This study presents the results of the time distribution as well as the cohort characteristics and their comparison. Furthermore, for convenience, not all analysis results have been presented; however, all metrics showed the exact same values, suggesting that the de-identification of the "source_value" fields does not affect the analysis.

We also reproduced the "MetforminVsSulfonylurea" study [25], which compares the risk of hypoglycemia among the users of metformin and sulfonylurea. Figure 2 shows the null distributions created in this study: The blue points in the figure represent the estimates of the negative control group on the log scale. Any estimates below the gray dashed lines (gray area) have a conventional P value <.05; the shaded orange area estimates have a calibrated P value <.05; and the pink area represents the 95% CI. In this case, every estimate from negative controls (dots) has a P value >.05 for both conventional and calibrated methods. In this analysis, although slightly different null distributions were generated for every iteration, this study commonly exhibited an area under the curve value of 1 with or without de-identification. Furthermore, we observed that the "source_concept_id" fields were used instead of the "source_value" fields in the SQL queries of this analysis.

Thus, these 2 experiments show that anonymization of the "source_value" fields can not only de-identify individuals in the CDM database but also simultaneously minimize the impacts of the de-identification on analysis.

Figure 1. Generated cohorts for "Influenza Cohort Diagnostics" study (A) without de-identification and (B) with the proposed de-identification strategy.

Cohort	jsh_InfluenzaHospCohortDiag	
	Entries	Subjects
Influenza narrow - with prior observation - 2009-2010 season	20	20
Influenza narrow - with no prior observation required - 2009-2010 season	63	63
Influenza broad - with prior observation - 2009-2010 season	75	75
Influenza broad - with no prior observation required - 2009-2010 season	195	195
Influenza narrow - with prior observation - 2014-2019 seasons	267	267
Influenza narrow - with no prior observation required - 2014-2019 seasons	444	444
Influenza broad - with prior observation - 2014-2019 seasons	1,704	1,704
Influenza broad - with no prior observation required - 2014-2019 seasons	2,992	2,992

(A)

Cohort	jsh_InfluenzaHospCohortDiag_no_source	
	Entries	Subjects
Influenza narrow - with prior observation - 2009-2010 season	20	20
Influenza narrow - with no prior observation required - 2009-2010 season	63	63
Influenza broad - with prior observation - 2009-2010 season	75	75
Influenza broad - with no prior observation required - 2009-2010 season	195	195
Influenza narrow - with prior observation - 2014-2019 seasons	267	267
Influenza narrow - with no prior observation required - 2014-2019 seasons	444	444
Influenza broad - with prior observation - 2014-2019 seasons	1,704	1,704
Influenza broad - with no prior observation required - 2014-2019 seasons	2,992	2,992

(B)

Figure 2. Null distributions for the "MetforminVsSulfonylurea" study (A) without de-identification and (B) with the proposed de-identification strategy. ITT: intention-to-treat.
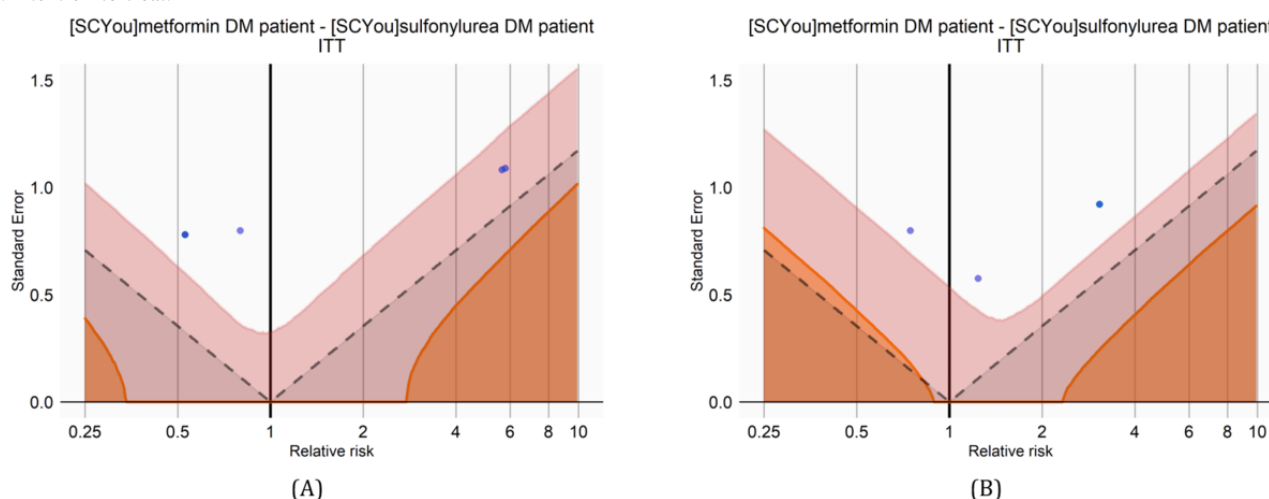
Discussion

Principal Findings

The OMOP-CDM has already implemented a high level of anonymization through the ETL process. However, it is possible to re-identify an individual. Therefore, when analyzing the CDM database in a cloud-computing environment or other public spaces, additional personal information de-identification is required.

The results of applying the proposed de-identification strategy along with the k-anonymity, l-diversity, and t-closeness privacy models to each table, particularly to the "source value" fields of the OMOP-CDM database, indicate that the strongest anonymity could be achieved with the t-closeness model. Moreover, the l-diversity and t-closeness models have stronger anonymization criteria than k-anonymity; these models increase the size of the groups using the same identifier and quasi-identifier values to achieve their criteria. However, while the k-anonymity model alone effectively prevents re-identification in the CDM to some extent, the l-diversity and t-closeness models perform better in terms of protecting personal information.

Finally, considering the other databases, the size of the CDM database increases as the operating period increases. Therefore,

de-identification of personal information should be periodically evaluated. Moreover, it is desirable to explore an appropriate privacy model and optimal conditions that suit the model.

Conclusions

Although the OMOP-CDM has no identifier nor foreign key that can be linked to the institutional database during the ETL process, a risk of personal information exposure remains because it preserves some source values. Therefore, we proposed a de-identification strategy that establishes transformation rules (see Table 1) for privacy models such as k-anonymity, l-diversity, and t-closeness, for the OMOP-CDM schema; this strategy complies with the recommended security policy. For instance, if the "person_source_value" of the PERSON table is "1234567890," it could be masked as "12345*****" after applying our de-identification strategy. This provides the flexibility to maintain the intentionally stored raw data. However, if the raw data exhibit a significantly low level of de-identification, it would be reasonable to not use the fields or tables. As a result of the evaluation of our de-identification strategy applied to the CDM database, it is possible to identify practical considerations for appropriate de-identification actions for each field. Thus, this research is a first step toward the development of safer and more appropriate de-identification policies for the OMOP-CDM schema and is expected to lay the foundation for further acceleration of CDM research.

Authors' Contributions

SJ designed the study. SJ, JS, and SK anonymized and measured the re-identification risks between the privacy models. JL, JK, and JS curated the data and supported the experiments. JM and JH equally contributed to this study as co-corresponding authors.

Conflicts of Interest

None declared.

References

1. Observational health data sciences and informatics. 2020. URL: https://www.ohdsi.org/ [accessed 2020-11-19]
2. Data-driven common data. PCORNET. 2020. URL: https://pcornet.org/data-driven-common-model/ [accessed 2020-11-19]
3. Sentinel common data model. Sentinel Initiative. 2020. URL: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model [accessed 2020-11-19]
4. ATHENA standardized vocabularies. Observational Health Data Sciences and Informatics. 2020. URL: https://www.ohdsi.org/analytic-tools/athena-standardized-vocabularies/ [accessed 2020-11-19]
5. ATLAS-A unified interface for the OHDSI tools. Observational Health Data Sciences and Informatics. 2020. URL: https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/ [accessed 2020-11-19]
6. ACHILLES for data characterization. Observational Health Data Sciences and Informatics. 2020. URL: https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/ [accessed 2020-11-19]
7. Fleisher RM. The Health Insurance Portability and Accountability Act. In: From Waiting Room to Courtroom: How Doctors can Avoid Getting Sued. Noida, India: JAYPEE; Jan 1, 2016.
8. Reiner Benaim A, Almog R, Gorelik Y, Hochberg I, Nassar L, Mashiach T, et al. Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies. JMIR Med Inform 2020 Feb 20;8(2):e16492 [FREE Full text] [doi: 10.2196/16492] [Medline: 32130148]
9. Rumbold J, Pierscionek B. Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix. JMIR Med Inform 2018 Nov 22;6(4):e47 [FREE Full text] [doi: 10.2196/medinform.7096] [Medline: 30467101]
10. Jonas S, Siewert S, Spreckelsen C. Privacy-Preserving Record Grouping and Consent Management Based on a Public-Private Key Signature Scheme: Theoretical Analysis and Feasibility Study. J Med Internet Res 2019 Apr 12;21(4):e12300 [FREE Full text] [doi: 10.2196/12300] [Medline: 30977738]
11. Griebel L, Prokosch H, Köpcke F, Toddenroth D, Christoph J, Leb I, et al. A scoping review of cloud computing in healthcare. BMC Med Inform Decis Mak 2015 Mar 19;15:17 [FREE Full text] [doi: 10.1186/s12911-015-0145-7] [Medline: 25888747]
12. OHDSI/ETL-CDMBuilder. GitHub. 2020. URL: https://github.com/OHDSI/ETL-CDMBuilder [accessed 2020-11-19]
13. OHDSI/USAGI. GitHub. 2020. URL: https://github.com/OHDSI/Usagi [accessed 2020-11-19]
14. OHDSI/DataQualityDashboard. GitHub. 2020. URL: https://github.com/OHDSI/DataQualityDashboard [accessed 2020-11-19]
15. OMOP Common Data Model. Observational Health Data Sciences and Informatics. 2020. URL: https://www.ohdsi.org/data-standardization/the-common-data-model/ [accessed 2020-11-19]
16. Prasser F, Kohlmayer F. Putting Statistical Disclosure Control into Practice: The ARX Data Anonymization Tool. In: Gkoulalas-Divanis A, Loukides G, editors. Medical Data Privacy Handbook. Cham, Switzerland: SpringerNature; 2015:111-148.
17. El Emam K. Guide to the De-Identification of Personal Health Information. Boca Raton, FL: Auerbach Publications; 2013.
18. OHDSI/OHDSIonAWS. GitHub. 2020. URL: https://github.com/OHDSI/OHDSIonAWS [accessed 2020-11-19]
19. Herold R, Beaver K. The Practical Guide to HIPAA Privacy and Security Compliance, 2nd edition. Boca Raton, FL: Auerbach Publications; Oct 20, 2014.
20. Sweeney L. k-Anonymity: a Model for Protecting Privacy. Int. J. Unc. Fuzz. Knowl. Based Syst 2012 May 02;10(05):557-570. [doi: 10.1142/S0218488502001648]
21. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 2007 Mar 01;1(1):3-52. [doi: 10.1145/1217299.1217302]
22. Li N, Li T, Venkatasubramanian S. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. 2007 Presented at: IEEE 23rd International Conference on Data Engineering; April 15-20, 2007; Istanbul, Turkey. [doi: 10.1109/ICDE.2007.367856]
23. Office for Government Policy Coordination (OPC), Ministry of the Interior Safety (MOIS), Korea Communications Commission (KCC), Financial Services Commission (FSC). Guidelines for de-identification of personal information. 2016 Jun 30. URL: https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000830764&fileSn=0 [accessed 2020-11-19]
24. ohdsi-studies/Covid19HospitalizationCharacterization/Influenza Cohort Diagnostics/. GitHub. 2020. URL: https://github.com/ohdsi-studies/Covid19HospitalizationCharacterization/tree/master/Influenza%20Cohort%20Diagnostics [accessed 2020-11-19]
25. ABMI / MetforminVsSulfonylurea. GitHub. 2020. URL: https://github.com/ABMI/MetforminVsSulfonylurea [accessed 2020-11-19]
26. ohdsi-studies/Covid19HospitalizationCharacterization. GitHub. 2020. URL: https://github.com/ohdsi-studies/Covid19HospitalizationCharacterization [accessed 2020-11-19]

Abbreviations

DEA: Drug Enforcement Administration
ETL: extract-transform-load
HIPAA: Health Information Portability and Accountability Act

NLP: natural language processing
NPI: National Provider Identifier
OHDSI: Observational Health Data Sciences and Informatics
OMOP-CDM: Observational Medical Outcomes Partnership Common Data Model

---

XSL•FO
RenderX

Original Paper

Effects of a Combination of Three-Dimensional Virtual Reality and Hands-on Horticultural Therapy on Institutionalized Older Adults' Physical and Mental Health: Quasi-Experimental Design

Tsung-Yi Lin[1], PhD; Chiu-Mieh Huang[2], PhD; Hsiao-Pei Hsu[2], MSN; Jung-Yu Liao[3], PhD; Vivian Ya-Wen Cheng[4], MEd; Shih-Wen Wang[5], MEd; Jong-Long Guo[5], PhD

[1]Department of Marketing and Distribution Management, Hsing Wu University, Taipei, Taiwan

[2]Institute of Clinical Nursing, School of Nursing, National Yang-Ming University, Taipei, Taiwan

[3]Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan

[4]PureAroma Healing Academy, Taipei, Taiwan

[5]Department of Health Promotion and Health Education, College of Education, National Taiwan Normal University, Taipei, Taiwan

Corresponding Author:
Jong-Long Guo, PhD
Department of Health Promotion and Health Education
College of Education
National Taiwan Normal University
No.162, Sec. 1, He-ping East Road
Taipei, 10610
Taiwan
Phone: 886 77493705
Email: jonglong@ntnu.edu.tw

Abstract

Background: Institutionalized older adults have limited ability to engage in horticultural activities that can improve their physical and mental health.

Objective: This study explored the effects of a combination of 3D virtual reality and horticultural therapy on institutionalized older adults' physical and mental health.

Methods: The study used a quasi-experimental design. A total of 106 older adults from 2 long-term care facilities were recruited and assigned to the experimental (n=59) or control (n=47) group. The experimental participants received a 9-week intervention. Both groups completed 3 assessments: at baseline, after the intervention, and 2 months later. The outcome variables included health status, meaning in life, perceived mattering, loneliness, and depression.

Results: The experimental group demonstrated significantly improved health status ($P<.001$), meaning in life ($P<.001$), and perceived mattering ($P<.001$) as well as significantly reduced depression ($P<.001$) and loneliness ($P<.001$) compared to the control group immediately after the intervention; these effects persisted for up to 2 months.

Conclusions: This study verified the beneficial effects of a combination of 3D virtual reality and hands-on horticultural therapy on older adults' health. These results could support the future successful implementation of similar programs for institutionalized older adults on a larger scale.

KEYWORDS

horticultural therapy; 3D VR; older adults; long-term care facility; mental health

Introduction

Caring for an aging global population is a critical issue facing countries throughout the world. As technological advances have improved the living conditions and health care of people, the population of older adults has risen sharply. According to a United Nations report [1], there were more than 901 million people over 60 years old in 2015 and only 607 million in 2000, which represents a 48% increase in the worldwide older adult population. The report also estimated that, by 2030, the world's

older adult population will reach 1.4 billion, and nearly 2.1 billion by 2050 [1]. Among older adults, gardening is a popular leisure-time activity, and exposure to nature has been found to produce positive effects and reduce psychological distress [2]. This may result from the natural environment's capacity to facilitate physical and psychological recovery by reducing fatigue and stress [3]. Therefore, there is emerging research that indicates that horticultural therapy can have a positive effect on human health, especially for older adults [4].

In previous studies [5,6], participants who received horticultural therapy demonstrated positive psychological, social, and physical health benefits. Horticultural therapy allows participants to work with plants and related products, and it promotes physical activity, which can provide physical benefits [7]. Another study [8] indicated that horticultural therapy could be an appropriate program for dementia care to serve older adults with cognitive, physical, and social needs. The horticultural therapy process allows participants to care for plants and perform related activities, thus increasing their physical, mental, and social well-being [9].

A previous comprehensive literature review [10] examined the effectiveness of gardening programs for both community-dwelling and institutionalized older adults. The review [10] included 22 studies with various research designs and indicated that gardening could promote overall health and quality of life, physical strength, fitness and flexibility, cognitive ability, and socialization. Although older adults can benefit from horticultural therapy activities, horticultural therapy currently requires participants to operate various types of gardening tools, which can be challenging. In addition, the preparation of flowers and plants can be complex and tedious, and instructors must pay attention to safety issues to avoid collisions, falls, and injuries. If an older adult could practice before performing gardening activities, this might be an effective way to avoid accidents and save workforce and material resources.

In the past, it was difficult to provide older adults with an opportunity to practice before horticultural therapy. However, the emergence of 3D virtual reality solves this problem. 3D virtual reality has gradually become more popular and has been widely used in various fields in recent years [11]. Virtual reality is a realistic virtual environment formed by a combination of computer software and hardware. Burdea [12] proposed that virtual reality should include 3 critical characteristics, namely, immersion, interaction, and imagination, which can make operators feel as if they are in the real world where they can interact instantly. For example, researchers found that after using virtual reality devices and a Nintendo Wii to perform cognitive training with older adult participants who exhibited mild cognitive impairment and dementia, there was a statistically significant effect on participants' overall cognition and visuospatial skills [13]. A virtual reality-based screening tool for cognitive function targeting older adults in primary care was found to be effective, and the participants expressed a positive perception and attitude toward virtual reality [14]. Another study [15] used virtual reality with hand-made devices for therapy to treat symptoms of apathy for the residents of long-term care facilities and provided evidence that it was feasible to use 3D

virtual reality during the intervention. Therefore, virtual reality may help older participants to overcome the difficulties experienced while engaging in physical gardening activities.

Also, 3D virtual reality educational activities can prevent overcrowding during horticultural therapy while increasing social participation and interpersonal communication through designed activities. Furthermore, as it is not limited by weather, place, or time, it is convenient for the residents of long-term care facilities to participate in. A prior study [16] compared life satisfaction and mood in a sample of 138 cognitively intact and ambulatory older adults, including 70 who lived in nursing homes, and 68 who lived independently in the community. Nursing home residents scored higher on the Depression-Dejection, Tension-Anxiety, and Confusion-Bewilderment subscales of the Profile of Mood State [16]. These findings show the importance of providing effective programs to lessen or alleviate unhealthy conditions for residents of long-term care facilities. Horticultural therapy has been found to be beneficial for health [17]. If new technological tools can be integrated with horticultural therapy and reduce the burden of human and material resources, this would represent a significant contribution to the literature.

Although the overall benefits of horticultural therapy have been established [18], the effectiveness of the combination of 3D virtual reality and hands-on horticultural therapy should be assessed. Therefore, this study aimed to explore this issue by considering several health-related outcomes of residents of a long-term care facilities.

Methods

Participants

A quasi-experimental design was used in this study (Clinicaltrials.gov; NCT04324203). One long-term care facility (>300 beds) out of the 6 long-term care facilities in North Taiwan was selected as the experimental group; another long-term care facility with similar characteristics and size was selected as the control group.

Participants were recruited through posters and verbal advertisements made by the staff of each long-term care facility during internal activities. The participants in this study were all long-term care facility residents who were over 65 years old. The selection criteria included being a long-term resident of the selected long-term care facilities, possessing the ability to understand verbal meanings, and being able to freely operate a joystick. The exclusion criteria included a history of severe psychiatric conditions, dementia, significant visual or hearing impairment, or current severe illnesses such as stroke or Parkinson disease.

Recruitment, Enrollment, and Assessments

Figure 1 shows a flowchart describing the enrollment and assessment process. After selecting the long-term care facility as the experimental group, the research team approached the executive director and staff to explain the purpose of the research and the methods and procedures that would be used in the study. After obtaining the agreement of the long-term care facility administration, we delivered recruiting messages

door-to-door and invited potential participants who met the inclusion criteria to participate in this study. The research team members also provided face-to-face explanations to potential participants, and after each participant completed the consent form, the research team members collected baseline data one by one in a quiet room provided by each long-term care facility.

Figure 1. Flowchart of participant enrollment and assessment.
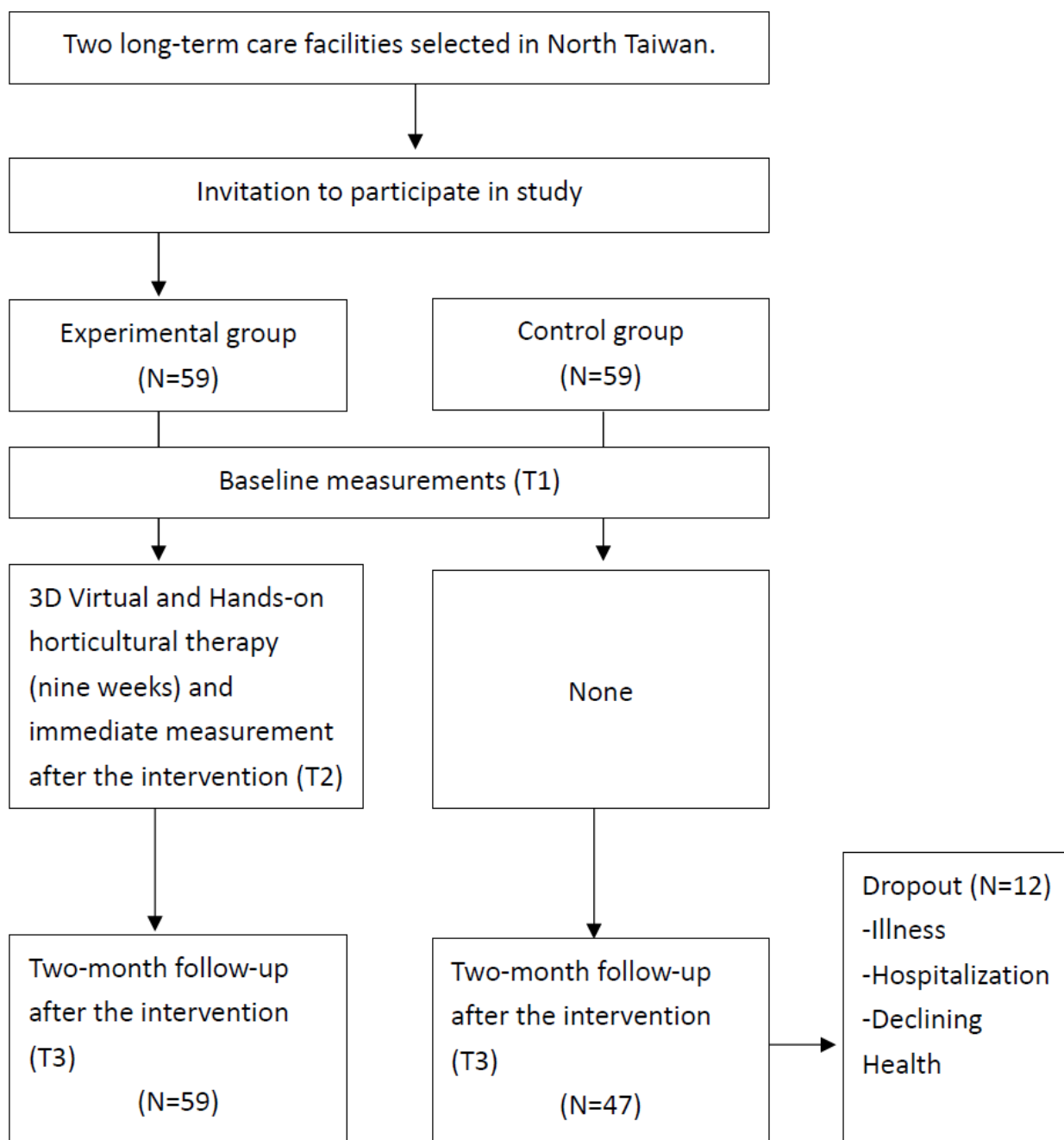
During the implementation period, medical professionals, long-term care facility staff, and horticultural therapy professionals were all present to ensure the safety of the participants. Members of the control group did not participate in any similar program during the intervention and follow-up periods.

The Combination of 3D Virtual and Hands-on Horticultural Therapy
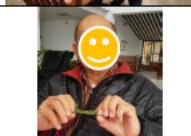
The intervention consisted of 18 one-hour sessions that occurred twice a week for 9 consecutive weeks. The first week introduced the participants to horticultural therapy and involved relationship-building activities such as knowing each other and remembering names as well as advising the participants on how to wear the 3D virtual reality helmets, operating the virtual reality handles, and familiarizing themselves with the virtual reality scenes. To avoid dizziness, the participants were allowed to practice using the virtual reality devices multiple times. The subsequent 8 sessions (developed by the research team) consisted of horticultural therapy, long-term care, older adults' health, and nursing professionals. The program components were developed based on the literature [19], with a focus on cultural competence. The 3D virtual reality and hands-on

horticultural therapy programs are described in Multimedia Appendix 1. Some examples of the 3D virtual reality and hands-on horticultural therapy programs are showed in Figure 2. Each week, there were 2 one-hour sessions that were facilitated by 2 certified horticultural therapy professionals and additional trained graduate students. During the intervention period, participants were allowed to use virtual reality–related equipment to perform additional exercises, with the assistance of the institutional staff. The cost information and details of the equipment and materials are showed in Multimedia Appendices 2-3

Figure 2. Example of the 3D virtual reality and hands-on horticultural therapy program and hand-made horticultural products. VR: virtual reality.

The intervention was designed to promote health status, meaning in life, and perceived mattering and also to reduce depression and loneliness. The participants were assigned to groups and worked with a facilitator who provided enhanced individual engagement and solved any problems with the operation of the 3D virtual reality device. The participants worked with the same group members and facilitator throughout the intervention. Graduate students were trained during a 16-hour 2-day workshop, in which they acquired basic knowledge of horticultural therapy skills so that they could assist the participants as they completed the activities.

We provided 3 sets of 3D virtual reality equipment. Participants took turns wearing the 3D virtual reality equipment in the same session. Each 3D virtual reality horticultural module lasted for 5 to 10 minutes. Although the time allotted for the first session was only 2 hours, the actual time spent in the session was over 3 hours. However, from the second session onward, as the participants became proficient in operating the 3D equipment,

the session time was reduced. The last session took approximately 2 hours to complete. In addition, we left the 3D equipment in the intervention long-term care facility to ensure that each participant could wear the equipment again if they wanted.

Measurement Instruments

The 12-item Chinese Health Questionnaire (CHQ-12) was used to evaluate the health of participants, and other outcome measures, such as meaning in life, perceived mattering, loneliness, and depression. Previous research [20] indicated that the use of the Chinese Health Questionnaire in older adults is as valid as in the general population. Meaning in life was associated with participants' multilevel health. For example, Pinquart [21] found that meaning in life was positively associated with physical health, social interaction, and interpersonal relationships among older adults. Elliot et al [22] defined perceived mattering as "the degree of being valued by significant others." It can be described as a kind of inner feeling that represents a 2-way interaction; the individual needs significant others to provide them with attention. A prior study [23] on loneliness found that loneliness could predict depression for up to 3 years. In addition, depression is one of the most common mental health problems among older adults [24]. While scoring each of the outcome variables, we reverse scored items with negative statements; thus, the directionality of all variables was consistent, and a higher number of points correlated with a better state of health for the participants.

To assess the health status of older adult participants, we used the 12-item Chinese Health Scale, which was developed by Cheng and Williams in 1986 [25] by translating the Goldberg General Health Questionnaire into Chinese. This scale consists of 12 Likert-type items that are scored from 3 (not at all) to 0 (more than usual). A sample item is "Feeling a headache or a sense of stress on the head?" The higher the score, the higher the level of general health. The Cronbach α coefficient was 0.83 in this study.

The meaning in life questionnaire was adapted from the Purpose in Life survey compiled by Frankl [26], and it consists of 9 Likert-type items. Each item is scored on a Likert-type scale from 1 (not at all) to 5 (very much), with higher scores indicating a higher level of perceived meaning in life. A sample item is "My life seems worthwhile." The Cronbach α coefficient was 0.87 in this study.

Perceived mattering was adapted from the General Mattering Scale [27] and consists of 5 items. Each item was scored on a Likert-type scale of 1 (not at all) to 4 (very much) with higher scores indicating a higher level of perceived mattering. A sample item is "How important do you feel you are to other people?" The Cronbach α coefficient was 0.86 in this study.

The short-form UCLA (University of California, Los Angeles) Loneliness Scale (ULS-6), which consists of 6 items [28], was used to measure loneliness. Each item was reverse scored on a Likert-type scale from 4 (never) to 1 (often), with higher scores indicating a lower level of perceived loneliness. A sample item is "I lack companionship." The Cronbach α coefficient was 0.83 in this study.

Using a Chinese version of the short-form of the Geriatric Depression Scale (GDS-15, [29]), we asked participants to describe their feelings over the prior week. A sample item was "Are you basically satisfied with your life?" All questions could be answered with 1 (yes) or 0 (no). The total raw score ranged from 0 to 15, with a higher score indicating a lower level of depression. The Cronbach α coefficient was 0.90 in this study.

The Chinese Health Questionnaire and the shortened Geriatric Depression Scale were available in Chinese. The other measurements included the Purpose in Life survey, General Mattering Scale, and short-form UCLA Loneliness Scale, which are available in English. We translated the 3 scales into Chinese and invited 6 professionals with expertise on Geriatrics and Gerontology to check content validity. It is suggested that the content validity index should be larger than 0.78 when content expertise is 6 or more [30]. In this study, all the scale items were larger than 0.80, with an average content validity index of 0.94.

Statistical Analyses

SPSS (version 20.0; IBM Corp) was used for the descriptive analyses of the demographic and outcome variables. Two-tailed independent t or chi-square tests were used to compare each difference (age, gender distribution, educational level, and economic status) between the experimental and control groups. The group comparisons of outcome measures at baseline were determined by performing Hotelling T^2 test to avoid type I errors. A generalized estimating equation was used to investigate the effects of time, group, and their interactions on the outcome variables; generalized estimating equations enable an understanding of change patterns over time at both the individual and group levels. A significance level of $P<.05$ was used.

Results

Participant Sociodemographic Data

The total number of participants who completed all activities was 106. Some participants (n=12) quit during the intervention because of illness, hospitalization, or declining health.

The participants' mean ages were 77.41 (SD 7.49) and 78.43 (SD 6.88) in the experimental and control groups, respectively. There were no statistically significant differences between groups in terms of the participants' age (P=.47), educational level (P=.052), or economic status (P=.28). However, there was a statistically significant difference in the gender distribution between the experimental and control groups. More men were in the intervention group than in the control group (intervention: 48/59, 81%; control: 25/47, 53%; P=.002). Because the gender distribution was significantly different, this confounding variable was controlled for in the generalized estimating equation analysis.

Improvements in Outcome Variables

The changes in each outcome variable over time are shown in Figure 3. The results of the generalized estimating equation analyses indicated that the members of the experimental group experienced significant improvements compared to their counterparts in the control group in terms of health status, meaning in life, perceived mattering, loneliness, and depression.

There was a significant group×time interaction for the 5 outcome measures. The experimental group demonstrated improvements in health status (β=9.99, *P*<.001), meaning in life (β=11.09, *P*<.001), perceived mattering (β=5.31, *P*<.001), loneliness (β=5.70, *P*<.001), and depression (β=4.49, *P*<.001).

Figure 3. Changes in outcome variables between the experimental group and the control group at 3 time points for (1) health status, (2) meaning in life, (3) perceived mattering, (4) loneliness (reverse scored), and (5) depression (reverse scored). T1: pretest; T2: postintervention; T3: 2-month follow-up.
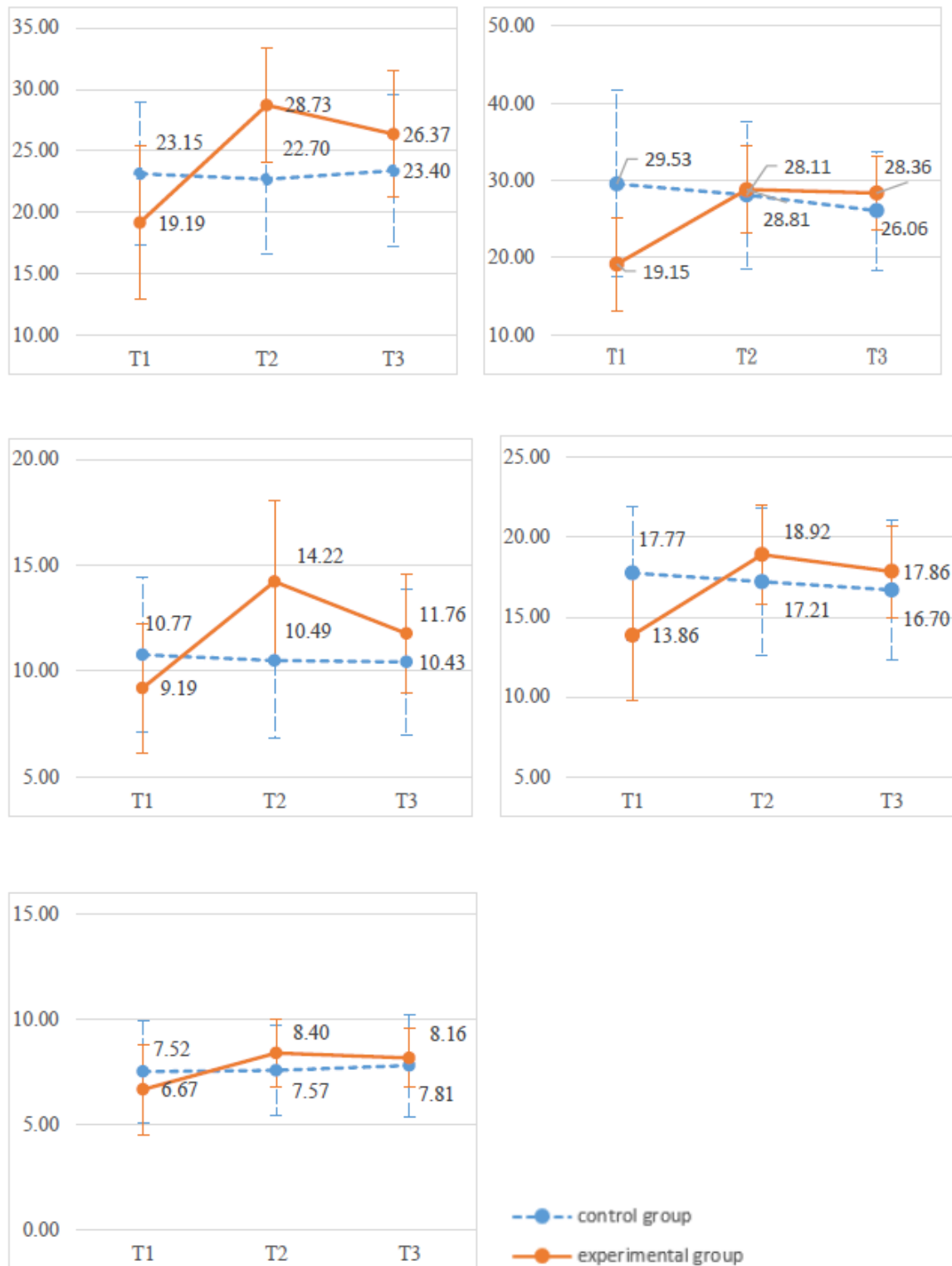
The results of the generalized estimating equation analyses also indicated that the experimental group exhibited significant effects when compared to the control group between the pretest and 2-month follow-up in terms of health status, meaning in life, perceived mattering, loneliness, and depression (Table 1).

The experimental group had improvements in health status (β=6.93, *P*<.001), meaning in life (β=12.67, *P*<.001), perceived mattering (β=2.92, *P*<.001), loneliness (β=5.13, *P*<.001), and depression (β=6.37, *P*<.001).

XSL•FO
RenderX

Table 1. Changes in outcome variables between the experimental group and the control group.

Outcome variable	β	SE	Wald χ^2	P value
Health status				
Group (experimental vs control)	–3.52	1.15	9.30	.002
Time (posttest vs pretest)	–0.45	0.27	2.83	.09
Time (2-month follow-up vs pretest)	0.26	0.58	0.19	.66
Group×time (experimental posttest vs control pretest)	9.99	0.94	112.41	<.001
Group×time (experimental 2-month follow-up vs control pretest)	6.93	1.00	47.99	<.001
Gender	1.57	1.07	2.16	.14
Meaning in life				
Group (experimental vs control)	–9.00	1.80	24.89	<.001
Time (posttest vs pretest)	–1.43	1.07	1.77	.18
Time (2-month follow-up vs pretest)	–3.47	1.33	6.84	.009
Group×time (experimental posttest vs control pretest)	11.09	1.42	60.89	<.001
Group×time (experimental 2-month follow-up vs control pretest)	12.67	1.55	66.95	<.001
Gender	4.88	1.50	10.59	.001
Perceived mattering				
Group (experimental vs control)	–1.20	0.67	3.18	.08
Time (posttest vs pretest)	–0.28	0.21	1.70	.19
Time (2-month follow-up vs pretest)	–0.34	0.52	0.43	.51
Group×time (experimental posttest vs control pretest)	5.31	0.67	63.56	<.001
Group×time (experimental 2-month follow-up vs control pretest)	2.92	0.70	17.31	<.001
Gender	1.35	0.58	5.38	.02
Loneliness				
Group (experimental vs control)	–3.79	0.79	0.59	.44
Time (posttest vs pretest)	–0.55	0.29	3.72	.054
Time (2-month follow-up vs pretest)	–1.06	0.43	6.18	.01
Group×time (experimental posttest vs control pretest)	5.70	0.65	77.93	<.001
Group×time (experimental 2-month follow-up vs control pretest)	5.13	0.72	51.43	<.001
Gender	0.74	0.75	0.99	.32
Depression				
Group (experimental vs control)	–2.72	0.80	11.48	.001
Time (posttest vs pretest)	0.18	0.43	0.18	.68
Time (2-month follow-up vs pretest)	–1.44	0.44	10.76	.001
Group×time (experimental posttest vs control pretest)	4.49	0.81	30.81	<.001
Group×time (experimental 2-month follow-up vs control pretest)	6.37	0.75	72.41	<.001
Gender	0.18	0.63	0.08	.78

Discussion

General

This interventional study was the first to combine 3D virtual reality with hands-on horticultural therapy and evaluate its impact on institutionalized older adults' physical and mental health. Elderly participants created hand-made horticultural products during weekly sessions and were asked to place the products in their rooms. One staff member working in the institution kept encouraging the participants to interact with the products. We believed that these strategies would increase the intervention intensity compared to those from strategies investigated in previous studies [9]. Previous studies may have lacked the supporting strategy and tracking design regarding the use of hand-made horticultural products [31]. The findings

of this study are in line with those from another study that combined 3D virtual reality and hands-on aromatherapy to improve institutionalized older adults' psychological health [32].

A previous study [33] indicated that the most common purpose of horticultural therapy was to improve the mood of participants, followed by social interaction, stress reduction, and motor skill development. A systematic review [19] of randomized controlled trials on the effectiveness of horticultural therapy also found significant effects on mental health and behavioral disorders, such as dementia, schizophrenia, and depression and on palliative-care for patients with cancer. Another systematic review [18] of the benefits of horticultural therapy for individuals with mental health conditions also supported the use of horticultural therapy for persons with one of mental health condition. These findings were in a variety of settings, for several mental health conditions, and with both males and females. The results of this study were consistent with those from prior research [6,18] and support the idea that horticultural therapy could be an effective program for improving residents' physical and mental health in long-term care facilities. It is important for the residents of long-term care facilities who are frail to interact with other people, for example interacting through the horticultural therapy program. The implementation of various aspects of horticultural activities as a health-promoting strategy can be used in long-term care facilities in both individual and group activities [10].

In a previous study [9] conducted in southern Taiwan to explore the effects of horticultural therapy on older adults in nursing homes, the experimental group received horticultural therapy for 1 hour once a week for 8 weeks, whereas the control group continued their routine daily activities. It was found that meaning in life was not improved in the experimental group; however, there were significant differences between the 2 groups on happiness and interpersonal intimacy. Compared to the findings of that study, our study found significant differences, both after the intervention and after 2 months, on meaning in life. The 3D virtual reality horticultural therapy program and the reminders provided by the long-term care facilities staff might both have contributed to the intensity of the intervention program; however, the isolated effects of each strategy need to be further explored in the future.

The participants informed the research team that they believed learning 3D virtual reality technology is beneficial for cognitive function; moreover, 3D virtual reality technology increased participants' learning motivation and allowed them to take the hand-made products with them to share with relatives and friends, thus increasing their interpersonal interactions. The horticultural therapists advised the participants to create the hand-made products as they are beneficial for physical and mental health. In the process of conducting the 3D and hands-on horticultural activities for participants, plenty of interpersonal interaction occurred between the researchers and the participants. For example, we found that the participants were unfamiliar with new 3D virtual reality technology and had to be patiently taught how to operate the technology by the researchers. At the beginning, the participants were apprehensive about operating the equipment. The researchers, therefore, used praise teaching to make participants feel confident to learn. The praise teaching method involves providing praise and encouragement to a participant whenever they complete a virtual reality operation or hand-made product to help them gain confidence to continue. Numerous participants reported looking forward to the weekly sessions because the sessions allowed participants to enjoy company, interact with others, and create hand-made products.

The improvements in perceived mattering that were observed indicated that the participants felt that the program gradually increased their value to their significant others, thereby improving their self-confidence and creating a sense of being treated seriously. During the intervention period, the researchers also found that after completing several sessions, older people were proud to share information with other residents regarding the techniques and methods of operating the 3D virtual reality equipment and demonstrating the finished products of their horticultural activities. This led other residents, who did not initially choose to participate in the experiment, to experience regret and ask the staff to participate in the middle of the intervention program.

A previous study [34] was conducted in central Taiwan on nursing home residents who participated in a weekly 1.5-hour-long indoor horticultural program that lasted for 10 weeks. This study [34] also found significant improvements in loneliness and depression, which were similar to the effects found in our study.

An extensive systematic review with meta-analysis [17] also found that horticultural activities had significant positive effects on a wide range of health outcomes, for example, reductions in depression and anxiety symptoms, stress, mood disturbances, and BMI, as well as increases in quality of life, sense of community, physical activity levels, and cognitive function. The researchers concluded that their meta-analysis provided robust evidence for the positive effects of horticultural therapy on health. Consistent with this meta-analysis [17], our results also indicate that the combined horticultural program could have a positive impact on residents of long-term care facilities. This research provides evidence that can be used by decision makers at long-term care facilities to add combined horticultural programs into their usual daily activities. For example, the director of the experimental institution expressed the intention to repeat a similar program annually to promote the health of the residents.

Limitations

Because the program integrated a combination of 3D virtual reality and hands-on horticultural therapy, the individual contribution of each approach could not easily be determined using this study's design. Therefore, to separately validate the effectiveness of the 2 approaches, additional studies should be conducted using controlled trials with sufficiently large sample sizes.

Another limitation was that we could not comment on the long-term effectiveness (ie, its effects 6 or 12 months later) of our intervention program. Additional follow-up is needed to

determine how the intervention program affects older adults' health long-term after completion of the intervention.

Conclusions

The results suggested that a 9-week program schedule might be sufficient to improve health status, meaning in life, perceived mattering, loneliness, and depression among residents of long-term care facilities both by the end of the intervention and 2-month later. Further studies are desirable to determine whether these improvements can be maintained with large-scale groups of participants and if the program produces long-term effects, such as decreases in the risk of mental health disorders

Conflicts of Interest

None declared.

Multimedia Appendix 1
Combination of 3D virtual reality and hands-on horticultural therapy: program components.
[DOC File , 46 KB - jmir_v22i11e19002_app1.doc]

Multimedia Appendix 2
Cost estimates of the 2 approaches.
[DOC File , 48 KB - jmir_v22i11e19002_app2.doc]

Multimedia Appendix 3
Information regarding equipment and materials.
[DOC File , 37 KB - jmir_v22i11e19002_app3.doc]

References

1. World population ageing report. United Nations. New York: United Nations Department of Economic and Social Affairs Population Division URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2020-02-15]
2. Hartig T, Evans GW, Jamner LD, Davis DS, Gärling T. Tracking restoration in natural and urban field settings. Journal of Environmental Psychology 2003 Jun;23(2):109-123. [doi: 10.1016/s0272-4944(02)00109-3]
3. Kaplan S. The restorative benefits of nature: Toward an integrative framework. Journal of Environmental Psychology 1995 Sep;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
4. Chan HY, Ho RC, Mahendran R, Ng KS, Tam WW, Rawtaer I, et al. Effects of horticultural therapy on elderly' health: protocol of a randomized controlled trial. BMC Geriatr 2017 Aug 29;17(1):192 [FREE Full text] [doi: 10.1186/s12877-017-0588-z] [Medline: 28851276]
5. Gigliotti CM, Jarrott SE, Yorgason J. Harvesting health: Effects of three types of horticultural therapy activities for persons with dementia. Dementia 2004;3(2):161-180. [doi: 10.1177/1471301204042335]
6. Han AR, Park SA, Ahn BE. Reduced stress and improved physical functional ability in elderly with mental health problems following a horticultural therapy program. Complement Ther Med 2018 Jun;38:19-23. [doi: 10.1016/j.ctim.2018.03.011] [Medline: 29857876]
7. Nicholas SO, Giang AT, Yap PL. The Effectiveness of Horticultural Therapy on Older Adults: A Systematic Review. J Am Med Dir Assoc 2019 Oct;20(10):1351.e1-1351.e11. [doi: 10.1016/j.jamda.2019.06.021] [Medline: 31402135]
8. Jarrott SE, Kwack HR, Relf D. An Observational Assessment of a Dementia-specific Horticultural Therapy Program. horttech 2002 Jan;12(3):403-410. [doi: 10.21273/horttech.12.3.403]
9. Yao YF, Chen KM. Effects of horticulture therapy on nursing home older adults in southern Taiwan. Qual Life Res 2017 Apr;26(4):1007-1014. [doi: 10.1007/s11136-016-1425-0] [Medline: 27696109]
10. Wang D, MacMillan T. The Benefits of Gardening for Older Adults: A Systematic Review of the Literature. Activities, Adaptation & Aging 2013 Apr;37(2):153-181. [doi: 10.1080/01924788.2013.784942]
11. Kim YM, Rhiu I, Yun MH. A systematic review of a virtual reality system from the perspective of user experience. Int J Hum Comput Interact 2019 Dec 13;36(10):893-910. [doi: 10.1080/10447318.2019.1699746]
12. Burdea G. Virtual reality systems and applications. 1993 Presented at: Electro'93 Intemational Conference; April 27-29; Edison, NJ.

XSL•FO
RenderX

13. Kim O, Pang Y, Kim JH. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]

14. Chua SIL, Tan NC, Wong WT, Allen Jr JC, Quah JHM, Malhotra R, et al. Virtual Reality for Screening of Cognitive Function in Older Persons: Comparative Study. J Med Internet Res 2019 Aug 01;21(8):e14821. [doi: 10.2196/14821]

15. Saredakis D, Keage HA, Corlis M, Loetscher T. Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed Methods Study. J Med Internet Res 2020 Jun 26;22(6):e17632 [FREE Full text] [doi: 10.2196/17632] [Medline: 32469314]

16. Gueldner SH, Loeb S, Morris D, Penrod J, Bramlett M, Johnston L, et al. A comparison of life satisfaction and mood in nursing home residents and community-dwelling elders. Archives of Psychiatric Nursing 2001 Oct;15(5):232-240. [doi: 10.1053/apnu.2001.27020]

17. Soga M, Gaston KJ, Yamaura Y. Gardening is beneficial for health: A meta-analysis. Prev Med Rep 2017 Mar;5:92-99 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.007] [Medline: 27981022]

18. Cipriani J, Benz A, Holmgren A, Kinter D, McGarry J, Rufino G. A systematic review of the effects of horticultural therapy on persons with mental health conditions. Occup Ther Ment Health 2017 Feb 07;33(1):47-69. [doi: 10.1080/0164212X.2016.1231602]

19. Kamioka H, Tsutani K, Yamada M, Park H, Okuizumi H, Honda T, et al. Effectiveness of horticultural therapy: a systematic review of randomized controlled trials. Complement Ther Med 2014 Oct;22(5):930-943. [doi: 10.1016/j.ctim.2014.08.009] [Medline: 25440385]

20. Chen CS, Tsang HY, Chong MY, Tang TC. Validation of the Chinese Health Questionnaire (CHQ-12) in community elders. Kaohsiung J Med Sci 2000 Nov;16(11):559-565. [Medline: 11294062]

21. Pinquart M. Creating and maintaining purpose in life in old age: A meta-analysis. Ageing Int 2002 Mar;27(2):90-114. [doi: 10.1007/s12126-002-1004-2]

22. Elliott G, Kao S, Grant AM. Mattering: Empirical Validation of a Social-Psychological Concept. Self and Identity 2004 Oct;3(4):339-354. [doi: 10.1080/13576500444000119]

23. Heinrich LM, Gullone E. The clinical significance of loneliness: a literature review. Clin Psychol Rev 2006 Oct;26(6):695-718. [doi: 10.1016/j.cpr.2006.04.002] [Medline: 16952717]

24. McCombe G, Fogarty F, Swan D, Hannigan A, Fealy GM, Kyne L, et al. Identified mental disorders in older adults in primary care: A cross-sectional database study. Eur J Gen Pract 2018 Dec;24(1):84-91 [FREE Full text] [doi: 10.1080/13814788.2017.1402884] [Medline: 29353511]

25. Cheng TA, Williams P. The design and development of a screening questionnaire (CHQ) for use in community studies of mental disorders in Taiwan. Psychol Med 1986 May;16(2):415-422. [doi: 10.1017/s0033291700009247] [Medline: 3726013]

26. Frankl VE. Man's search for meaning: An introduction to logotherapy. New York, NY: Pocket; 1963.

27. Marcus FM. Mattering: Its measurement and theoretical significance for social psychology. 1991 Presented at: Annual meeting of the Eastern Sociological Association; 1991; Cincinnati, OH.

28. Zhou L, Li Z, Hu M, Xiao S. Reliability and validity of ULS-8 loneliness scale in elderly samples in a rural community. J Cent South Univ 2012 Nov;37(11):1124-1128 [FREE Full text] [doi: 10.3969/j.issn.1672-7347.2012.11.008] [Medline: 23202622]

29. Cheng ST, Chan ACM. A brief version of the geriatric depression scale for the chinese. Psychol Assess 2004 Jun;16(2):182-186. [doi: 10.1037/1040-3590.16.2.182] [Medline: 15222814]

30. Waltz C. Measurement in nursing research. In: Measurement in nursing research. Philadelphia: F. A. Davis; 1991.

31. Park SA, Lee AY, Lee GJ, Kim DS, Kim WS, Shoemaker CA, et al. Horticultural Activity Interventions and Outcomes: A Review. Korean J. Hortic. Sci. Technol 2016 2;34(4):513-527. [doi: 10.12972/kjhst.20160053]

32. Pfeffer JC, Deyton DE, Fly JM. Survey of Horticultural Therapy Programs in Tennessee. Journal of Therapeutic Horticulture 2009;19:24-44.

33. Cheng VYW, Huang CM, Liao JY, Hsu HP, Wang SW, Huang SF, et al. Combination of 3-Dimensional Virtual Reality and Hands-On Aromatherapy in Improving Institutionalized Older Adults' Psychological Health: Quasi-Experimental Study. J Med Internet Res 2020 Jul 23;22(7):e17096 [FREE Full text] [doi: 10.2196/17096] [Medline: 32706660]

34. Chen YM, Ji JY. Effects of horticultural therapy on psychosocial health in older nursing home residents. Journal of Nursing Research 2015;23(3):167-171. [doi: 10.1097/jnr.0000000000000063]

XSL•FO
RenderX

Original Paper

Pharmacists' Perceptions of the Benefits and Challenges of Electronic Product Information System Implementation in Hong Kong: Mixed-Method Study

Eunice Wing To Fung[1], B (Pharm); Gordon Tsz Fung Au-Yeung[1], B (Pharm); Lo Mei Tsoi[2], B (Pharm); Lili Qu[2], B (Pharm); Tommy Kwan Wa Cheng[3], B (Pharm); Donald Wing-Kit Chong[3], B (Pharm), MBA; Teddy Tai Ning Lam[1], PharmD, PhD; Yin Ting Cheung[1], BSc (Pharm), PhD

[1]School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[2]Pfizer Corporation Hong Kong Limited, Hong Kong, China (Hong Kong)
[3]GlaxoSmithKline Consumer Healthcare (Hong Kong) Limited, Hong Kong, China (Hong Kong)

Corresponding Author:
Yin Ting Cheung, BSc (Pharm), PhD
School of Pharmacy, Faculty of Medicine
The Chinese University of Hong Kong
8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building, Area 39
Shatin, New Territory
Hong Kong
China (Hong Kong)
Phone: 852 39436833
Email: yinting.cheung@cuhk.edu.hk

Abstract

Background: With the advancement of technology, more countries are now adopting the use of electronic product information (ePI), which refer to an electronic version of physical product inserts in a semistructured format optimized for electronic manipulation. The successful implementation of ePI has led to advantages and convenience to patients, health care professionals, and pharmaceutical companies in many regions and countries. In the Hong Kong Special Administrative Region (SAR), there is currently no citywide implementation of ePI. The SAR exhibits conditions that would favor the implementation of an ePI system, as well as existing barriers hindering its implementation. However, no study has been performed to examine the specific situation in Hong Kong.

Objective: The objective of this study is to explore working pharmacists' overall perception of ePI and to identify potential challenges to the implementation of an ePI system in Hong Kong.

Methods: This mixed-method study involved a structured survey and interview with practicing pharmacists in Hong Kong. Pharmacists were eligible if they were licensed to practice in Hong Kong, and currently working locally in any pharmacy-related sectors and institutions. Respondents completed a survey to indicate their level of agreement with statements regarding the potential advantages of ePI over paper PI. A structured interview was conducted to gather respondents' perceived advantages of ePI over paper PI in different aspects, such as professionalism, usability, presentation, and environment, as well as challenges of citywide ePI implementation in Hong Kong. Thematic analysis was adopted to analyze the qualitative data. Grounded theory was used to generate themes and identify specific outcomes.

Results: A total of 16 pharmacists were recruited, comprising 4 community pharmacists, 5 hospital pharmacists, and 7 industrial pharmacists. All of them used electronic platforms at least once per month on average. Respondents identified many flaws in physical package inserts that can potentially be mitigated using ePI. The speed with which drug information can be retrieved and the degree to which the drug information can be readily updated and disseminated were considered the greatest strengths of ePI. The clarity with which ePI present drug information to patients was considered as the weakest aspect of ePI. Many respondents highlighted concerns about the security risks and high cost associated with system maintenance and that certain subpopulations may not be sufficiently computer literate to navigate the ePI system. Respondents also voiced many concerns about the implementation and maintenance of a local ePI system.

XSL•FO
RenderX

Conclusions: We conclude that an ePI system is generally supported by pharmacists but concerns about implementation process and maintenance of the system has been raised. The perceived benefits of ePI gathered from this study, as well as collective evidence from other countries with mature ePI systems, confirm that more efforts should be made to promote optimized development and implementation of an ePI system in Hong Kong.

Introduction

Product inserts in the packaging of medicines typically include a summary of product characteristics (SmPCs) and a package leaflet, and are regulated and legally required in most countries. The accelerating progress of technology has also led to the development of electronic product information (ePI) for medicines, and this is now gaining popularity in Australia, Canada, Switzerland, Japan, the United States, and countries within the European Union (EU) [1-5].

According to an international report published by the European Medicines Agency (EMA), most regions and countries regard ePI as product information for electronic handling. It is composed of an SmPC, a package leaflet, and product labeling in a semistructured format under a common electronic standard stipulated by governing bodies [2]. Common electronic standards comprise mark-up languages, standardized vocabularies, and internationally recognized interoperability specifications. Specifically, text is annotated to formats such as eXtensible Markup Language (XML), JavaScript Object Notation (JSON), and HyperText Markup Language (HTML), with XML being the most commonly adopted format. Standardized vocabularies are used to describe medicine information, to enable searches across different product information (PI). Interoperability specifications enable the integration of ePI into other health systems to facilitate its use by a wide range of health care professionals. Its semistructured format allows the inclusion of structured elements (fixed headings and controlled vocabularies) and unstructured elements (free text and graphics) to tailor product information to individual needs [2]. Most ePI formats can be disseminated via the web, e-platforms (software or websites and tools such as computers, mobile devices, and wearables), and print media.

Different studies of health care professionals' perceptions of ePI have been conducted in various countries. Ahead of the launch of an ePI pilot scheme in Belgium and Luxembourg, results of a survey of hospital pharmacists demonstrated strong support for the use of ePI for patient services [6]. Ninety percent of the respondents supported a complete switch from paper to electronic versions of product information, and up to 55% of the respondents used only electronic package "leaflets" in their daily practice. Another study examined pharmacists' readiness for paperless labeling in the United States [7], and a study assessing community pharmacists' perception of tailor-made electronic leaflets was also conducted in Portugal in 2018 [8]. The results of both studies showed that pharmacists had positive attitudes about the use of individually tailored patient leaflets

to address the individual needs of their patients, and many pharmacists believed that such initiatives could improve the health literacy of patients. In general, the results of studies performed in many countries have demonstrated that pharmacists strongly support the implementation of ePI systems.

In the Hong Kong Special Administrative Region (SAR), there is currently no citywide implementation of ePI. However, the SAR exhibits several conditions that would favor the implementation of an ePI system. For example, there is a clear set of laws in place to regulate the registration and labeling of pharmaceutical products in Hong Kong [9]. In addition, there is a high penetration rate of technology in Hong Kong: in 2018, over 90% of households had access to the internet at home [10]. Moreover, more than 90% of people aged 10 years or over reported that they had used the internet in the past 12 months to seek information that they needed.

Despite the near-ubiquitous use of the internet in Hong Kong households, online drug information searching is not commonly performed. Of the 443 respondents who participated in a local survey, less than half (44%) reported that they had looked for health information online, mostly using professional websites including government and hospital sites [11]. The majority of respondents surfed for disease-specific information and general information on healthy lifestyle. However, only 10.9% of respondents got online information on drugs and medications [11]. Respondents who were older, had lower than tertiary education, and had lower monthly income were less likely to surf online health information. Respondents also indicated that they were uncertain whether the information found online was reliable and accurate. Although this local study is limited by a convenience sampling approach, its finding suggests that there are indeed barriers hindering the implementation of ePI systems but research in this area is limited, and no other study has been performed to examine the specific situation in Hong Kong.

In Hong Kong, pharmacists play an important role in traditional drug distribution and dispensing duties. Over the recent years, a paradigm shift from mere product dispensing to more patient-oriented delivery of pharmacy service has been promoted [12]. Pharmacists are becoming more actively involved in making clinical decisions and patient care in both hospital and community settings. Hong Kong is also a hub for many international pharmaceutical companies. Most local country offices are focused on medical affairs and sales and marketing of pharmaceutical products in Asia. Other than practicing as authorized persons in this capacity, industrial pharmacists facilitate quality control, regulation, and registration of new products [12]. The execution of such services requires

pharmacists of different practicing sectors to apply their professional drug information retrieval skills from multiple sources, particularly from package inserts. Consequently, obtaining product information from centralized ePI systems may be more efficient than using paper PI. Moreover, ePI may be used as a supporting platform to allow pharmacists to conduct patient education activities. A locally adapted set of ePI may improve information sharing and communication between patients and pharmacists, thereby empowering patients in the management of their health.

The objectives of this study were to explore working pharmacists' overall perception of ePI, with a focus on gathering the perceived benefits and barriers that pharmacists face when they use ePI in their practice, and identifying potential challenges to the implementation of an ePI system in Hong Kong. It is expected that the findings of this study will reflect the acceptance level of pharmacists toward ePI and will yield insights into necessary improvements and objectives for future studies related to ePI.

Methods

Study Design

This was a mixed-method study that involved a structured questionnaire and individual interviews with practicing pharmacists in Hong Kong between January and May 2020. Approval from the Survey and Behavioral Ethics Committee of the Chinese University of Hong Kong was obtained before the initiation of this study (Reference number: SBRE-19-204). Written informed consent was obtained from all of the participants.

Study Population

Pharmacists were eligible if they were (1) licensed to practice in Hong Kong, (2) self-identified as capable of reading and writing in English (as most drug information platforms available in Hong Kong are in English), and (3) currently working locally in any pharmacy-related sectors and institutions, including but not limited to public and private hospitals, community pharmacies, pharmaceutical companies, distributors, government authorities, pharmacy-related nongovernmental organizations, and private clinics. Eligible pharmacists were identified from the above sectors and recruited purposively using a snowball sampling approach. As the survey and interview were both conducted in Chinese, respondents must be able to converse in Cantonese and read Traditional Chinese. Eligible pharmacists were first approached via email where the study objectives were briefly described, and upon informal consent, they were subsequently contacted by phone to arrange for a face-to-face interview.

Research that has evaluated sample-size requirements in qualitative studies shows that data saturation often occurs at approximately the 12th to 16th respondent in a homogenous group [13,14]. Therefore, the minimum target sample size in this study was 12 respondents. After the 12th participant, recruitment was to continue until data saturation was reached.

Study Procedures

Survey

After obtaining informed consent, the overarching aim of this research (ie, exploring the prospect of developing an ePI system in Hong Kong) was briefly described to pharmacists (respondents). They first completed a structured online questionnaire that was developed based on reported experience with ePI by other countries and regions [1-6], as well as consensus from research team comprising practicing pharmacists (LT, LQ, TC, and DC), senior pharmacy students (EF and GA-Y) and research methodologists (TL and YC). Functionality of the electronic survey was pilot tested on 2 research assistants. This was a closed survey where only respondents had access to the survey via a protected link. The survey consisted of 2 major sections.

The first section was a set of closed-ended questions to collect baseline characteristics, which included practicing sector and site and years of working experience. The second section consisted of questions to evaluate respondents' level of agreement with each statement about the potential advantages of ePI over paper PI in different aspects, such as professionalism, usability, presentation, and environment. Some examples of the statements were "Using ePI instead of PI can enhance communication with patients" and "Using ePI instead of PI can enhance adequacy of drug information" (Multimedia Appendix 1). The agreement score was presented in a Likert scale ranging from 1 (ePI is much worse than paper PI) to 10 (ePI is much better than paper PI). The questionnaire was developed and administered in Traditional Chinese, an official written language in Hong Kong. It was translated to English for reporting purposes (Multimedia Appendix 1).

Structured Interview

After completing the online questionnaire, respondents were interviewed individually. The interviewers (EF and GA-Y) did not have prior relationship or interaction with the respondents. Their experience with using ePI and their perceptions of the benefits and challenges of citywide ePI implementation in Hong Kong were discussed systematically using a structured script (Multimedia Appendix 2). Their opinions on the method of implementation and the feasibility of a complete replacement of paper PI by ePI were also collected. All of the interviews were conducted in Cantonese, an official spoken language in Hong Kong, and were audio recorded. Each interview lasted between 20 and 40 minutes, and was held privately at the workplace of the respondents.

Special Considerations During the COVID-19 Pandemic

Three interviews were completed in January 2020 during the pre-COVID-19 phase. The remaining 13 interviews were conducted between mid-April and May when the COVID-19 situation in Hong Kong had become relatively under control. Respondents were given a choice of whether they would like to have face-to-face or phone interviews, which all 13 respondents indicated that they preferred meeting in-person. The Prevention and Control of Disease (Prohibition on Group Gathering) Regulation (Cap. 599) came into operation on March

29, 2020 to prohibit any group gathering of more than 4 persons in any public place in Hong Kong [15]. Hence, each respondent completed the interview with only 2 investigators (EF and GA-Y). Basic social distancing measures (keeping a distance of at least 1 m from each other and wearing of facemasks) were strictly adhered to. In view of the ongoing pandemic, no repeat interview was conducted.

Data Analysis

Online survey responses were recorded anonymously and retrieved electronically from the online survey software Qualtrics (SAP). Only completed survey was analyzed. All of the descriptive data collected from the survey were analyzed using the Statistics Package for Social Science, version 25 (IBM Corp.). The CHERRIES Checklist [16] was adopted to report the online survey findings.

All of the interviews were transcribed verbatim in Chinese. Data were managed using Archiv für Technik, Lebenswelt und Alltagssprache (ATLAS).ti version 8 qualitative data-analysis software (Scientific Software Development GmbH). Thematic analysis was adopted to analyze the qualitative data. Grounded theory was used to generate relevant themes and identify specific outcomes. Grounded theory was chosen as the theoretical framework as it is one of the most widely used systematic approaches to analyze qualitative data in the field of social sciences and health care [17,18]. There are limited data in the literature on pharmacists' perception of ePI. The inductive nature of grounded theory is applied to obtain a deeper understanding of the rationale behind nascent barriers or perceived benefits of ePI that are unique to the local population.

Two researchers (EF and GA-Y) read all of the transcripts and coded them independently. Coding was conducted in 2 cycles: in the first cycle, codes were created and data segments were assigned to the codes; in the second cycle, the code lists were validated and then applied to the remaining data. Codes were then cross-checked and reviewed by a third researcher (YC). Thereafter, the coding and themes were discussed with the research team, and a coding framework was developed and applied to all of the transcripts. The number of recurring themes was noted accordingly. Finally, the themes and codes were translated into English for reporting purposes. The COREQ (COnsolidated criteria for REporting Qualitative research) 32-item checklist [19] was adopted to report the qualitative findings (Multimedia Appendix 3).

Results

Working Experience and Characteristics of Respondents

A total of 16 pharmacists were recruited, comprising 4 community pharmacists, 5 hospital pharmacists, and 7 industrial pharmacists (Table 1; response rate: 100%). The majority of pharmacists (respondents) reported having 1-10 years of working experience.

Almost all of the respondents considered themselves as regular retrievers of drug information (n=15/16, 94%), and all used electronic platforms at least once per month on average (n=16/16, 100%; Figure 1).

More than 85% (n=14/16, 88%) of the respondents commonly used package inserts or electronic platforms for retrieving information about dosages and methods of administration, and 75% (n=12/16) commonly retrieved information about indications, contraindications, and adverse effects of drugs (Figure 2).

Table 1. Characteristics of study population (N=16).

Characteristics	Number of respondents (%)
Current practicing sectors	
Hospital	5 (31)
Community pharmacy	4 (25)
Multinational pharmaceutical company	7 (44)
Previous practicing sectors[a]	
Public hospital	8 (50)
Private hospital	2 (13)
Community pharmacy (chain or independent)	5 (31)
Multinational pharmaceutical company	7 (44)
Nongovernmental organizations	2 (13)
Others	3 (19)
Years of working experience	
1-5	5 (31)
5-10	3 (19)
10-15	1 (6)
15-20	3 (19)
Did not mention	4 (25)
Most frequently used electronic devices for accessing online drug information	
Laptop	6 (38)
Smartphone	10 (63)

[a]Proportions do not add up to 100% as a pharmacist can have experience in multiple sectors.

Figure 1. Frequency of retrieving drug information from different platforms (N=16). Physical PI refers to the conventional paper product information within the package of the drug product. Other physical platforms include paper copies of MIMS, BNF, LexiComp Handbook etc. Electronic platforms include electronic product information systems (eg, eMC UK, DailyMed/FDA), drug databases (eg, Lexicomp, Micromedex, AccessPharmacy), point-of-care databases (eg, ClinicalKey, UpToDate, Dynamed Plus), and Pharmaceutical (prescribing) references (eg, electronic MIMS, BNF). BNF: British National Formulary; eMC: Electronic Medicines Compendium; FDA: Food and Drug Administration; MIMS: Monthly Index of Medical Specialities; PI: product insert.
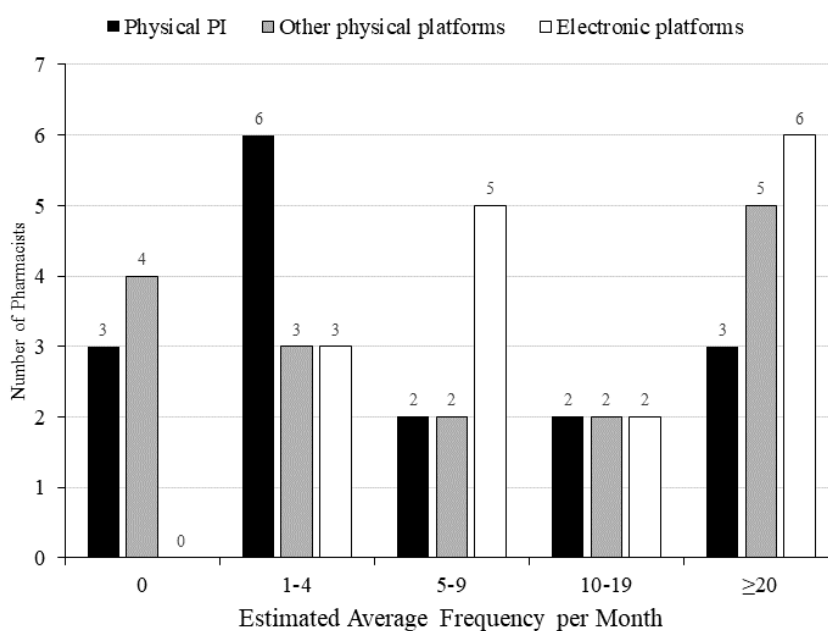
Figure 2. Categories of Drug Information Commonly Retrieved by Pharmacists (N = 16).
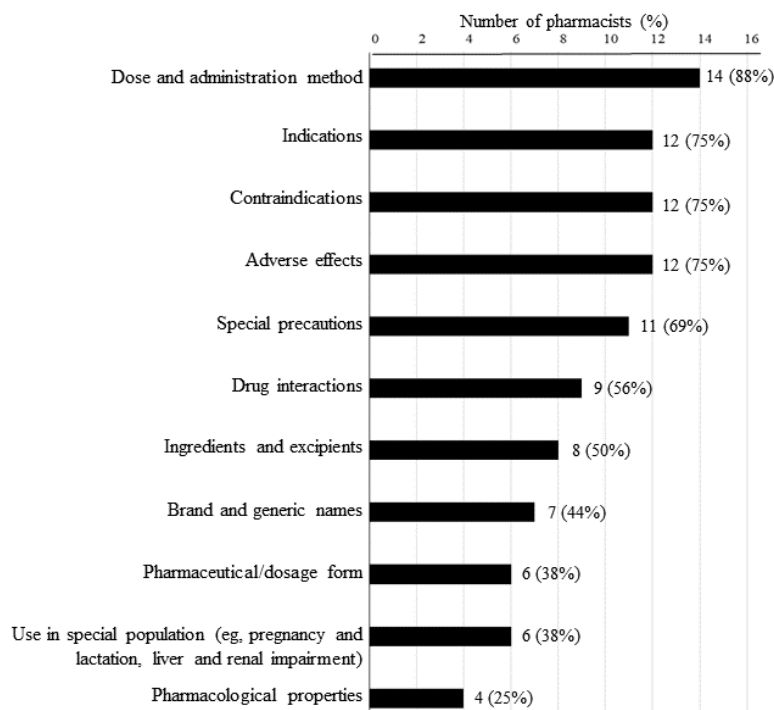
Survey

In general, respondents indicated that their experience with ePI had been entirely positive (Table 2). The speed with which drug information can be retrieved and the degree to which the drug information is up to date were considered the greatest strengths of ePI, being given an average score of 9.29 out of 10 by the respondents. The clarity with which ePI presents drug information to patients was considered as the weakest aspect of ePI, with an average score of only 6.93 out of 10.

Table 2. Pharmacists' perceived impact of electronic product information (ePI) system on different aspects of their work.

Impacts of ePI	Mean score[a] (SD)
Professionalism	
Patient safety	7.14 (1.79)
Communication with patients	6.93 (2.06)
Communication with prescribers	8.07 (1.33)
Usability	
Adequacy of drug information	7.93 (1.73)
Retrieving speed	9.29 (0.73)
Accessibility of drug information	8.64 (1.34)
Degree to which the drug information is up to date	9.29 (0.83)
Presentation	
Organization and layout	7.57 (1.74)
Environment	
Impact on the environment	9.14 (0.77)
Overall	
Overall impact	8.00 (1.30)

[a]The agreement score range was from 1 (ePI is much worse than paper PI) to 10 (ePI is much better than paper PI).

Structured Interview

Data saturation was reached and no new coding was observed after 15 interviews. Five main themes were observed in the result, namely, the problem of paper package inserts and current electronic drug information sources (Theme 1), the advantages of ePI (Theme 2), the challenges of ePI implementation (Theme 3), concerns about the replacement of paper PI with ePI (Theme

4), and suggestions on how to implement a local ePI system in Hong Kong (Theme 5).

Theme 1: Problems With Paper Package Inserts and Current Electronic Drug Information Sources

Most respondents agreed that there are many obvious problems with paper package inserts, especially in their layout. In particular, it was noted that the font size on package inserts is too small, making the insert difficult to read:

> *The layouts of package inserts designed by different companies can be very different, and the fonts used are so small that I always have to spend a long time finding the information I want.* [Respondent 01, industrial pharmacist]

Another respondent stated an example:

> *When doctors call and ask questions, we usually have to answer within a few minutes, and it is too time-consuming to look for answers on package inserts.* [Respondent 02, industrial pharmacist]

Furthermore, 2 respondents stated that sometimes they cannot retrieve information from a package insert if they do not have the drug on hand or if the drug is in a sealed package. That is, respondents are not able to retrieve information from package inserts at their own convenience. In this situation, respondents indicated that they look for information on other electronic sources, via a search engine or electronic databases. The most frequently mentioned databases were the Monthly Index of Medical Specialities (MIMS) (n=9/16, 56%), Lexicomp (n=7/16, 44%), and Micromedex (n=6/16, 38%). However, some respondents stated that the information on MIMS is too general and that they cannot find the information they want. Another respondent pointed out there may be out-of-date or even wrong information on the MIMS website or in search engine results:

> *There is just general information on MIMS, such as storage conditions. There is not enough detailed information, such as clinical trial data. Some people may prefer a complete PI over MIMS.* [Respondent 03, industrial pharmacist]

Theme 2: Perceived Advantages of ePI

Overall, respondents noted that ePI had either user-related or technical advantages (Table 3). The most noted benefits of ePI were improved access to the most updated information and the reduced cost of printing paper leaflets.

Table 3. Major advantages and disadvantages of electronic product information identified by respondents.a

Advantages	Disadvantages
User related	
• Facilitate retrieval of the most updated product labeling information	• Need for patients to adapt to assess information electronically
• Improved patient access to product information	• Patients' preference for paper leaflet within the drug package instead of browsing a website for drug information
• Open access to information	• Accessibility of information is limited as it is technology dependent
• Allow prompt communication of new safety information to health care professional/public	• Not easy to highlight important points
• Facilitate alignment of information among different brands of the same drug	
Technical related	
• Reduce printing cost/redressing cost of package leaflet	• Retrieval of information is not possible when there is no internet access
• Allow prompt implementation of new indication	• Risk of security issues or cyber attack
• More environmentally friendly	• Need for tight control and annotations over the different versions of electronic product information
	• High cost for development and maintenance of a reliable platform

aInterview transcripts were reviewed. Major themes on advantages and disadvantages of electronic product information were identified using thematic analysis.

Respondents considered the ease of retrieving drug information as the biggest advantage of ePI, as it enables them to do so without having the physical package inserts on hand.

> *The keyword searching function makes it more convenient to retrieve the information we want.* [Respondent 04, industrial pharmacist]

Respondents who worked in public hospitals stated that each hospital has a standard practice of scanning package inserts and

uploading them to a centralized database to facilitate retrieval of drug information by doctors and pharmacists. They postulated that it would be more convenient if a formal ePI system were established. Another advantage identified was that ePI facilitates the retrieval of up-to-date information. For example, amendments or new information on the drug product can be annotated, updated, and disseminated electronically immediately after approval from regulatory authorities; this is a more timely and efficient approach, as compared to reprinting, repackaging, and distributing the revised paper PI. The update of new information can also be communicated to the user through a notification system or "pop up message."

A key advantage of ePI for industrial pharmacists is being able to view the updated version of a package insert directly on a website, without any other procedures being required. For hospital pharmacists and community pharmacists, a key advantage of ePI is being able to ensure that the information they are viewing is up to date. Finally, 1 respondent pointed out that implementing an ePI system is more environmentally friendly than printing paper-package inserts.

Theme 3: Challenges of ePI Implementation

Despite the many advantages of ePI mentioned during the interviews, many respondents highlighted concerns about the security risks and high cost associated with system maintenance and that certain subpopulations may not be sufficiently computer literate to navigate the ePI system (Table 3).

Respondents mentioned that their biggest concern with the implementation of a local ePI system is that there may be a legal problem; specifically, that placing package inserts on a website may violate Cap.231 *Undesirable Medical Advertisements Ordinance* (UMAO), which states that no person shall publish, or cause to be published, any advertisement likely to lead to the use of any medicine for the purpose of treating human beings for any disease and conditions specified [10]. Examples of such conditions include gynecological or obstetrical disease, correction of deformity or the surgical alteration of a person's appearance, and treatment of benign or malignant tumor. Thus, it will depend on the manner in which drug information is placed online. For example, a paper PI that is inside any container or package containing an oral drug does not constitute "publication of advertisement." However, if a Quick Response (QR) code was printed on the outside of a package such that consumers could access the information by scanning the QR code, some respondents suggested that this may potentially violate the law. One respondent suggested that

> *It is illegal for a pharmaceutical company to publish information that claims a product has curative or preventive effects on cancer to be open to the public. Such locally published information may violate UMAO* [Respondent 05, industrial pharmacist]

Another challenge of implementation that was identified in interviews is that there is a low incentive for implementation of an ePI system in Hong Kong. One respondent explained that given the fact that drug information can be obtained from various existing sources, he does not expect governing authorities to take the initiative in developing a local ePI system.

As compared to establishing an online drug information system, the government might allocate its resources to more pressing and ongoing "e-health projects," such as setting up the electronic health record sharing system between private and public sectors, and the Chinese Medicine information system. Respondents were also skeptical on whether the pharmaceutical industry in Hong Kong is supportive of this initiative, though the 5 industrial pharmacists interviewed in this study (representing different pharmaceutical companies) seemed to agree with the rationale and anticipated benefits behind establishing an ePI system. In addition, the public's potential level of acceptance of an ePI system is unknown, and local data on support for ePI initiatives are lacking.

Other difficulties in implementation were suggested to be the initiation and maintenance of the system. Some respondents thought that it will be difficult to include all drugs in a single system, as there are many registered pharmaceutical products in Hong Kong.

> *There are many registered brands for certain drugs in Hong Kong. Some community pharmacies may sell different brands of the same drug, and we notice that the package inserts of different brands can be slightly different.* [Respondent 06, community pharmacist]

Therefore, this respondent thought that the large number of registered pharmaceutical products from different companies may increase the difficulty of managing the labeling content of all drugs in an ePI system. For example, the potential challenges in the standardization of the ePI format were also mentioned, although standardization would enable faster retrieval of information. There is a need to establish systems and processes to enforce compliance with labeling and formatting standards; this may likely be cost intensive. Therefore, many respondents were concerned about the maintenance of the ePI system; 1 respondent pointed out that more resources may be needed for system maintenance:

> *If the information is updated more frequently, the Drug Office [the law enforcement agency that governs all legislations concerning medicines in Hong Kong] will have to review more changes of particulars, and more human resources may be needed.* [Respondent 07, hospital pharmacist]

The costs associated with implementation and system maintenance are issues of concern. For example, one respondent mentioned that the transition from paper PI to ePI itself might pose a major logistic burden to some pharmaceutical companies:

> *We have to convert the existing PI into ePI format. This is a significant challenge for the industry. We will probably need high-end IT support. If a company, like mine, has many registered products, this will require significant resources. For the smaller companies, they would still need to invest resources on ePI compliant systems.* [Respondent 10, industrial pharmacist]

Theme 4: Concerns With Complete Replacement of Paper PI With ePI

Some respondents were concerned about the accessibility of information, especially for the elderly population and patients with low computer literacy:

> *As we can see in hospitals, most of the patients are elderly, and they may not know how to access the information from the Internet.* [Respondent 08, hospital pharmacist]

Another respondent also shared a similar thought:

> *Many elderly people do not know how to use electronic devices, and they may rush to a community pharmacy and ask for help to retrieve online drug information, increasing the burden on us as community pharmacists.* [Respondent 06, community pharmacist]

The same respondent also expressed concern over what might happen in a situation of system breakdown:

> *If we rely totally on an online ePI system, in the situation of a system breakdown we will not have any other alternative ways to retrieve drug information.* [Respondent 06, community pharmacist]

Overall, most respondents suggested that an ePI system should be implemented and that the physical package insert system should also be maintained.

Theme 5: Enablers for Implementing a Citywide ePI System in Hong Kong

The majority agreed that a local ePI system should be initiated and managed by regulatory sectors in Hong Kong. A few respondents suggested that a centralized platform should be made for pharmaceutical companies to upload the latest version of their products' ePI, and that the authority should take responsibility for verifying the data and ensuring their accuracy before releasing the information to the public.

> *The authority should set up some guidelines for the requirements of ePI format and information, so that pharmaceutical companies can know what to include and what to exclude in the drug information they upload onto an ePI system.* [Respondent 09, industrial pharmacist]

Most respondents also suggested that there should be separate information for both health care professionals and the public within the same ePI system. For health care professionals, detailed prescribing information and product-specific information should be provided. For the public, abridged patient-friendly information and nontechnical language should be used. Some respondents also suggested that patient-information leaflets in drug packaging inserts should be translated into Chinese, as they thought that a simplified patient-information leaflet in Chinese language would attract more patients to read the drug information:

> *The public will not be interested to read a long and complicated package insert, so a simpler version should be tailor-made for the public.* [Respondent 10, industrial pharmacist]

Finally, several features that can be integrated into an ePI system were mentioned, such as adding images of the pill and package. One respondent also suggested adding audio and video versions of drug information. Another suggested adding a *drug interactions* check function and *drug compatibility* check function:

> *Compatibility of a drug with other IV fluids is a common question, but sometimes information about compatibility is not included in the package insert, so it should be added to ePI system if possible.* [Respondent 11, hospital pharmacist]

Discussion

Principal Findings

In this study, we sought to determine pharmacists' (our respondents') perceptions of the benefits and challenges of implementing an ePI system in Hong Kong. Our respondents showed that pharmacists generally have a positive attitude toward ePI. Specifically, they stated that although they always use package inserts to retrieve drug information, they found many flaws in physical package inserts that can potentially be mitigated using ePI. However, they had mixed views on the replacement of paper PI by ePI; the majority thought that paper PI should be kept, rather than replacing with ePI. Respondents also voiced many concerns about the implementation and maintenance of a local ePI system. Most stated that further discussion and coordination of different stakeholders are needed before initiating a citywide ePI system in Hong Kong.

Our study shows that the majority of respondents reported frequent use of electronic databases to retrieve drug information, such as MIMS, Lexicomp, and Micromedex. However, several problems related to their use were identified. Currently, no study has systematically evaluated local pharmacists' perception of existing online drug databases. However, anecdotal finding from our respondents, as well as reports from other countries, suggested that the common problems identified by health care professionals were the lack of comprehensive coverage of drug items, incomplete information (eg, missing information on drug compatibility for intravenous administration), and underutilization of internet features that enhance readability and understanding of medication information (eg, font enlargement, glossary of terms) [20,21]. As identified from a local study that reviewed local drug information use, authors found that despite the widespread use of the internet in Hong Kong, the public is still not benefiting from existing online drug information databases [22]. Authors identified that this was attributable to the readability of existing online drug information, as well as a language barrier, as most of the information is provided in English rather than Chinese [22]. The implementation of an ePI system is expected to ameliorate these problems, as it will offer a platform with comprehensive drug information in a standardized format for health care professionals, as well as patient-friendly information in Chinese for the public.

Several advantages of ePI were identified in this study. The ease of retrieving drug information is considered to be greatly enhanced by ePI, as users can access drug information whenever they need it, even when they do not have the drug package on-hand. The experience of other countries with ePI confirms that increased availability of drug information can be enabled by an ePI system [2,6,7]. In addition, online drug information can be constantly updated, which will assist the work of health care professionals in reducing drug-related problems in a hectic clinical setting. This is especially relevant to the context of Hong Kong health care system, where staff shortages, workforce problems, and timeliness of services are key shortcomings [12,23]. It is expected that a mature ePI system would serve as a one-stop drug information hub to help improve delivery of information to users.

Despite its many advantages, one of the challenges identified in the interviews with respondents was that an ePI may violate Cap.231 *UMAO*. According to Cap.231, any notice, poster, circular, label, wrapper, or document is considered to be an advertisement. The law states that no person shall publish, or cause to be published, any advertisement likely to lead to the use of any medicine for the purpose of treating human beings for any disease and conditions specified [24]. It is important to note that this law is targeted at advertisements that are published locally within the jurisdiction of Hong Kong. An ePI is technically not an "advertisement" by nature. The responders raised this point probably because they held a strict interpretation of the ordinance that ePI published online locally might be deemed as an "advertisement" under UMAO, and therefore a contravention of the ordinance. We also identified 2 exemptions concerning this ordinance. First, there are exemptions for providing health care professionals with drug information; the law also states that "it shall be a defense to prove that the advertisement to which the proceedings relate was made only in a publication of a technical character intended for circulation" mainly among medical practitioners, pharmacists, and the medical and para-medical staff of hospitals, clinics, and maternity homes [10]. This implies that the UMAO will not be violated if ePI system access is limited to health care professionals. Second, if open access to the public is desired, discussion with the government will be required. An amendment of the law may be necessary to exempt product information that is publicly available on the local ePI system from being referred as "advertisement." Resources will also be required to manage the large database of drug information and to verify the information provided by manufacturers.

Although the respondents showed support for the implementation of a local ePI system, the majority wanted paper PI to be kept to serve as an alternative for special circumstances, such as during a system breakdown or other technical problems. In particular, respondents expressed concerns over the ability of the elderly population and patients with low computer literacy to use ePI. According to a study about online health information-seeking behavior in Hong Kong, a "digital divide" may exist, such that senior citizens and individuals of a lower socioeconomic status and a lower education level may be less likely to have access to online drug information [11]. This would be a problem, as Hong Kong has a growing elderly population who may have limited access to digital information. Accordingly, we propose that local champions should be identified to serve as advocates to help deliver and promote the benefits of ePI. Our findings also reveal the need for ePI-related education to be incorporated, starting from the undergraduate level to enable professional development and empower practitioners to effectively utilize ePI for drug-information retrieval.

In summarizing all recommendations raised by respondents, we have identified 2 viable approaches to establishing a centralized ePI system. The first approach is to take advantage of the current databases and drug websites that are managed by the local authorities, onto which the Drug Office can undergo transition into an XML format of PI, and followed by pharmaceutical companies during registration or during changes of the particulars of pharmaceutical products. Providing drug information in XML format should become a mandatory requirement after the successful transition. The second approach is to develop a new centralized platform where pharmaceutical companies submit drug information and labeling content of their pharmaceutical products according to predefined standards in XML format. Aside from being the developer and regulator of the ePI system, the government could also provide financial support for pharmaceutical companies to transit into the ePI system. It is also possible that pharmaceutical companies can collaborate and develop a consortium or consensus on the key principles, common standards, and governance process of the ePI system. The Hong Kong Association of the Pharmaceutical Industry may facilitate such initiative, together with other stakeholders including consumers, health care professionals, academic institutions, and regulatory authorities.

The collective opinions and recommendation gathered from the respondents suggest that an integrated drug information system should be implemented to serve both professional users and the public. Therefore, we propose that Hong Kong could adopt a similar platform as the Electronic Medicines Compendium (eMC), which is a well-developed ePI system used in the UK. eMC is managed by a third-party provider, and pharmaceutical companies pay to upload their SmPC and a package leaflet onto the eMC website [1]. There is a standardized format and layout for drug information on the website, with in-depth information on product characteristics supplied for health care professionals, and an abridged version available for patients and the public. The UK authorities also play a pivotal role in the management of this ePI system, as all drug information uploaded on the website must be checked and approved by the UK/European medicines licensing authority (Medicines and Healthcare products Regulatory Agency [MHRA]/EMA). Currently, the eMC has more than 10,600 documents that have been verified and approved by the UK or European government agencies that license medicines. This system is well-received by the citizens of the UK and EU, as well as other parts of the world. Hong Kong could benefit from adopting key features of the eMC system, as well as integration of other reputable and complementary sources of non-drug–related health information. For example, the system can include portal links to the public health care system website (Hospital Authority [25]), government websites (Hong Kong Department of Health [26],

Centre for Health Protection [27], Chinese Medicine Regulatory Office [28]), or clinical trial portals and databases on adverse drug reactions. It is anticipated that this ePI system can be a leading, authoritative platform for health and drug information on the internet in Hong Kong.

Limitations of Study

The study population comprised pharmacists only, who are presumably the health care professionals that most clinicians and the public would consult for drug-related inquiries. More studies are required to understand the acceptance levels of other health care professionals for ePI, including doctors and nurses, as they are also frequent users of package inserts. In addition, the sample size in this study was small. However, saturation was identified and no new coding was observed after 15 interviews. We also acknowledge that the data generated in the interviews broadly revolve around the needs of pharmacists, patients, and policy makers (eg, the government), with minimal references to other health care professionals. This might be due to the nature of the questions asked in the structured interviews, as there was not sufficient contact time to ask respondents to discuss other specific issues related to ePI. Given the nascent status of the ePI field in Hong Kong, we reason that despite these limitations, there is still reasonable credibility of these preliminary qualitative findings in demonstrating potential relevance of an ePI system in Hong Kong.

Study participants were recruited using purposive snowball sampling, and this may have resulted in selection bias. Although the respondents reported experience in multiple sectors of practices, only pharmacists who were currently practicing in hospital, community, and industrial settings participated in the study. Notably, these are also the 3 major sectors of practice for 80% of the pharmacists in Hong Kong [29]. We speculate that the identified benefits and challenges are generally reflective of the current opinions of local pharmacists. Lastly, there are methodological limitations associated with grounded theory approach. For example, the investigators conducted an extensive literature review on ePI before the initiation of this study. This might lead to the prejudgment or formation of predetermined constructs even before data analysis. To address this limitation, at least two investigators conducted coding independently to facilitate fine-tuning of the analysis and converge on a shared interpretation.

Given the aforementioned limitations of this study, we would like to emphasize that this study should be interpreted as an effort to generate future research directions in this field. We hope to take advantage of existing models in other countries and identify essential features of well-established and effective ePI systems. A follow-up study should then evaluate how these features can be adapted to serve the needs of health care professionals and patients in Hong Kong. Future studies should also seek the opinions and expertise of different stakeholders on ways to maintain an ePI system to ensure its sustainability. In particular, the attitudes of other health care professionals and patients toward electronic drug information should also be evaluated when new services are introduced and when they are more adapted to reading the electronic versions of patient drug-information leaflets.

Conclusion

The establishment of a citywide system of ePI would be in keeping with the vision embodied by "The Smart City Blueprint for Hong Kong" [30], which was introduced by the Government of the Hong Kong SAR to facilitate the use of innovation and technology to improve people's quality of living.

This qualitative study is the first step in exploring the feasibility of implementing a citywide ePI system in Hong Kong. Our findings suggest that pharmacists generally have a positive attitude toward ePI implementation, as they highlighted the advantages of ePI in enabling better updating and retrieval of drug information while also recognizing the many challenges that need to be overcome. In particular, Hong Kong pharmacists generally appear to not be in favor of a total replacement of paper PI by ePI. Nevertheless, the perceived benefits of ePI gathered from this study, as well as collective evidence from other countries with mature ePI systems, confirm that more efforts should be made to promote optimized development and implementation of an ePI system in Hong Kong.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Perceived Impact of Electronic Product Information (ePI) on Pharmacists' Work (Questionnaire).
[DOCX File , 20 KB - jmir_v22i11e20765_app1.docx]

Multimedia Appendix 2
Structured Interviewer Guide.
[DOCX File , 16 KB - jmir_v22i11e20765_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

CHERRIES Checklist.
[DOCX File , 26 KB - jmir_v22i11e20765_app3.docx]

References

1. Knight N, Moreland J. Digitising product information: Greater accessibility for public health & all patients. URL: https://www.ema.europa.eu/en/documents/presentation/presentation-digitising-product-information-greater-accessibility-public-health-all-patients-n_en.pdf [accessed 2020-05-11]

2. European Medicines Agency, Heads of Medicines Agencies, European Commission. Electronic product information for medicines in the EU: Report from an EMA–HMA–EC workshop held on 28 November 2018. 2019. URL: https://www.ema.europa.eu/en/documents/report/report-european-medicines-agency-ema/heads-medicines-agencies-hma/european-commission-ec-workshop-electronic-product-information-epi_en.pdf [accessed 2020-05-11]

3. United States National Library of Medicine. DailyMed. 2020. URL: https://dailymed.nlm.nih.gov/dailymed/index.cfm [accessed 2020-06-17]

4. Australian Government Department of Health. Therapeutic Good Administration. 2020. Product Information. URL: https://www.tga.gov.au/product-information-0 [accessed 2020-06-17]

5. Pharmaceuticals and Medical Devices Agency (PMDA) - Japan. Safety Alert & Recalls/Review Reports/Package Inserts. URL: https://www.pmda.go.jp/english/search_index.html [accessed 2020-06-17]

6. Lambot N. E-PIL Belgium & Luxembourg pioneer pilot project. URL: https://www.ema.europa.eu/en/documents/presentation/presentation-e-pil-belgium-luxembourg-pioneer-pilot-project-n-lambot_en.pdf [accessed 2020-05-11]

7. Ho Y, Chen Q, Nian H, Johnson KB. An assessment of pharmacists' readiness for paperless labeling: a national survey. J Am Med Inform Assoc 2014;21(1):43-48 [FREE Full text] [doi: 10.1136/amiajnl-2013-001654] [Medline: 23523874]

8. Vitor C, Cavaco A. Community pharmacists' attitudes towards patient leaflets: Exploring perceptions underlying an electronic local production of tailored written information. AIMS Public Health 2018;5(2):189-202 [FREE Full text] [doi: 10.3934/publichealth.2018.2.189] [Medline: 30094280]

9. Drug Office, Hong Kong. Guidance notes on registration of pharmaceutical products/substances. 2019. URL: https://www.drugoffice.gov.hk/eps/do/en/doc/guidelines_forms/guid.pdf [accessed 2020-05-11]

10. Census and Statistics Department. Information technology usage and penetration. Thematic Household Survey Report No. 67. 2019. URL: https://www.statistics.gov.hk/pub/B11302672019XXXXB0100.pdf [accessed 2020-05-11]

11. Yan YY. Online health information seeking behavior in Hong Kong: an exploratory study. J Med Syst 2010 Apr;34(2):147-153. [Medline: 20433053]

12. Lee CP. Health Care System and Pharmacy Practice in Hong Kong. Can J Hosp Pharm 2018;71(2):140-148 [FREE Full text] [Medline: 29736048]

13. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Nov 21;18(1):148. [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]

14. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822x05279903]

15. Hong Kong e-Legislation. Cap 599A. Prevention and Control of Disease Regulation. 2020. URL: https://www.elegislation.gov.hk/hk/cap599A [accessed 2020-06-10]

16. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

17. Chapman AL, Hadfield M, Chapman CJ. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. J R Coll Physicians Edinb 2015;45(3):201-205. [doi: 10.4997/JRCPE.2015.305] [Medline: 26517098]

18. Foley G, Timonen V. Using Grounded Theory Method to Capture and Analyze Health Care Experiences. Health Serv Res 2015 Aug;50(4):1195-1210 [FREE Full text] [doi: 10.1111/1475-6773.12275] [Medline: 25523315]

19. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

20. Raban MZ, Tariq A, Richardson L, Byrne M, Robinson M, Li L, et al. Evaluation of Web-Based Consumer Medication Information: Content and Usability of 4 Australian Websites. Interact J Med Res 2016 Jul 21;5(3):e21 [FREE Full text] [doi: 10.2196/ijmr.5651] [Medline: 27443680]

21. Ghoshal M, Walji MF. Quality of medication information available on retail pharmacy Web sites. Res Social Adm Pharm 2006 Dec;2(4):479-498. [doi: 10.1016/j.sapharm.2006.05.002] [Medline: 17161807]

22. Cheng FWT, So SWK, Fung BWT, Hung WH, Lee VWY. Online drug information platform for the public in Hong Kong-Review of local drug information use and needs. Int J Med Inform 2018 Jun;114:27-34. [doi: 10.1016/j.ijmedinf.2018.03.006] [Medline: 29673600]

23. Schoeb V. Healthcare Service in Hong Kong and its Challenges. China Perspectives 2016 Dec 01;2016(4):51-58 [FREE Full text] [doi: 10.4000/chinaperspectives.7118]

XSL•FO
RenderX

24.   Hong Kong e-Legislation. Cap 231. Undesirable Medical Advertisements Ordinance. 2020. URL: https://www.elegislation.gov.hk/hk/cap231 [accessed 2020-05-11]
25.   Hospital Authority. URL: https://www.ha.org.hk/visitor/ha_index.asp [accessed 2020-06-10]
26.   The Government of Hong Kong SAR. Department of Health. URL: https://www.dh.gov.hk/eindex.html [accessed 2020-06-10]
27.   The Government of Hong Kong SAR. Department of Health. Centre for Health Protection. URL: https://chp.gov.hk/en/index.html [accessed 2020-06-10]
28.   The Government of Hong Kong SAR. Department of Health. Chinese Medicine Regulatory Office. URL: https://www.cmro.gov.hk/html/eng/index.html [accessed 2020-06-10]
29.   The Government of Hong Kong SAR (Department of Health). 2015 Health Manpower Survey. Summary of the Characteristics of Pharmacists Enumerated. 2015. URL: https://www.dh.gov.hk/english/statistics/statistics_hms/sumpharm15.html [accessed 2020-06-10]
30.   The Government of the Hong Kong SAR (Office of the Government Chief Information Office). Hong Kong Smart City Blueprint. URL: https://www.smartcity.gov.hk/ [accessed 2020-06-17]

Abbreviations

ATLAS: Archiv für Technik, Lebenswelt und Alltagssprache
COREQ: COnsolidated criteria for REporting Qualitative research
EMA: European Medicines Agency
eMC: Electronic Medicines Compendium
ePI: electronic product information
EU: European Union
HTML: HyperText Markup Language
JSON: JavaScript Object Notation
MHRA: Medicines and Healthcare products Regulatory Agency
MIMS: Monthly Index of Medical Specialities
QR: Quick Response
SAR: Special Administrative Region
SmPCs: summary of product characteristics
UMAO: Undesirable Medical Advertisements Ordinance
XML: eXtensible Markup Language

XSL•FO

RenderX

Original Paper

An Environmental Scan of Sex and Gender in Electronic Health Records: Analysis of Public Information Sources

Francis Lau[1*], PhD; Marcy Antonio[1*], BSc, MPH; Kelly Davison[1*], MSc, MN; Roz Queen[1*], BA; Katie Bryski[2], BA

[1]School of Health Information Science, University of Victoria, Victoria, BC, Canada

[2]Canada Health Infoway, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Francis Lau, PhD
School of Health Information Science
University of Victoria
P.O. Box 1700 STN CSC
University of Victoria
Victoria, BC, V8W2Y2
Canada
Phone: 1 2504725131
Email: fylau@uvic.ca

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e30764/

Abstract

Background: Historically, the terms sex and gender have been used interchangeably as a binary attribute to describe a person as male or female, even though there is growing recognition that sex and gender are distinct concepts. The lack of sex and gender delineation in electronic health records (EHRs) may be perpetuating the inequities experienced by the transgender and gender nonbinary (TGNB) populations.

Objective: This study aims to conduct an environmental scan to understand how sex and gender are defined and implemented in existing Canadian EHRs and current international health information standards.

Methods: We examined public information sources on sex and gender definitions in existing Canadian EHRs and international standards communities. Definitions refer to data element names, code systems, and value sets in the descriptions of EHRs and standards. The study was built on an earlier environment scan by Canada Health Infoway, supplemented with sex and gender definitions from international standards communities. For the analysis, we examined the definitions for clarity, consistency, and accuracy. We also received feedback from a virtual community interested in sex-gender EHR issues.

Results: The information sources consisted of public website descriptions of 52 databases and 55 data standards from 12 Canadian entities and 10 standards communities. There are variations in the definition and implementation of sex and gender in Canadian EHRs and international health information standards. There is a lack of clarity in some sex and gender concepts. There is inconsistency in the data element names, code systems, and value sets used to represent sex and gender concepts across EHRs. The appropriateness and adequacy of some value options are questioned as our societal understanding of sexual health evolves. Outdated value options raise concerns about current EHRs supporting the provision of culturally competent, safe, and affirmative health care. The limited options also perpetuate the inequities faced by the TGNB populations. The expanded sex and gender definitions from leading Canadian organizations and international standards communities have brought challenges in how to migrate these definitions into existing EHRs. We proposed 6 high-level actions, which are to articulate the need for this work, reach consensus on sex and gender concepts, reach consensus on expanded sex and gender definitions in EHRs, develop a coordinated action plan, embrace EHR change from socio-organizational and technical aspects to ensure success, and demonstrate the benefits in tangible terms.

Conclusions: There are variations in sex and gender concepts across Canadian EHRs and the health information standards that support them. Although there are efforts to modernize sex and gender concept definitions, we need decisive and coordinated

actions to ensure clarity, consistency, and competency in the definition and implementation of sex and gender concepts in EHRs. This work has implications for addressing the inequities of TGNB populations in Canada.

Introduction

Background

The collection of sex and gender data as part of a person's demographic information in electronic health records (EHRs) has been in place for decades [1]. Historically, the 2 terms have been used interchangeably as a binary attribute to describe an individual as being male or female. There has been little effort to define this attribute and its values consistently across EHRs, health care organizations, and health information standards [2]. Longstanding use of poorly understood and defined sex and gender concepts has created much systemic inconsistencies. Different EHRs use either sex or gender as the data element name for this attribute and a wide range of coding schemes with letters and numbers such as *F*, *M*, *1*, *2*, *U*, and *UNK* to represent the values [2-4]. There is also growing recognition that sex and gender are distinct concepts [5]. Sex is a biological construct based on anatomy, genetics, and hormones that distinguish between males and females on a continuum. Gender is a psychological and social construct based on attitudes, feelings, behaviors, and cultural factors a person may use to identify and express their authentic identity as a man, woman, or other options. Given the diversity that exists in our society, it is inaccurate, inadequate, and inappropriate to use a single binary attribute to represent sex and gender of the transgender and gender nonbinary (TGNB) populations [6,7].

The inability to capture accurate sex and gender data in EHRs has also created a structural barrier to the health of the TGNB populations [2]. Without accurate sex and gender documentation, health care staff are unable to identify TGNB patients properly, which can lead to distress, stigma, and delay in seeking care for these patients [6,7]. The estimated TGNB populations in the United States range from 0.5% to 4.5% among adults and 2.5% to 8.4% among children and adolescents with upward trends noted, but the actual proportions can vary as many remain invisible in EHRs [8]. Jeffee et al [9] reported that 1 in 4 TGNB patients seeking health care services were denied equal treatment, and over 40% of transgender men experienced verbal harassment, physical assault, or denial of care. Studies have shown TGNB individuals to have worse health care access, quality, and outcomes, including lower rates of health maintenance screening and life expectancies, and higher rates of depression, substance use, and chronic illness [2,9,10]. For instance, in a study of 5135 US veterans diagnosed with gender dysphoria, the odds were 4 times greater for depression and 5 times greater for HIV seropositivity compared with controls [10]. The mismatch of one's sex and gender can also be problematic with sex-based care guidelines such as pregnancy and prostate cancer screening tests. For transgender patients who have changed their recorded birth sex to match their current

gender, the built-in EHR rules would often reject these tests for not having the expected sex value present [11-13].

There have been efforts to improve the definition, collection, and use of sex and gender data in EHRs among health care organizations. For instance, in 2013 the World Professional Association for Transgender Health published a set of recommendations for EHR developers, vendors, and users with respect to transgender patients [14]. In the United States, the Centers for Medicare and Medicaid Services and Office of the National Coordinator (ONC) for health information technology (IT) require EHR vendors to include sex and gender data fields as part of the EHR software certification [15], but health care providers are currently not required to collect this information [16]. Notable examples where gender identity is collected and used include the Fenway Institute [17] and the US Department of Veterans Affairs [11]. However, there are wide variations in current practices [8,18,19], and many health care organizations are yet to implement a system to collect sex and gender data [2]. The lack of common standards in defining sex and gender data fields and code values has also made it difficult to exchange and reuse this information across EHRs [20]. Overall, this is a complex multi-level challenge with numerous implementation barriers to explore for health policy, health systems, IT, providers, researchers, and patients.

Objectives

The purpose of this environmental scan is to understand how sex and gender data are defined and implemented in existing Canadian EHR systems and current international health information standards. This study informs a larger initiative underway by the authors to modernize sex and gender information practices in Canadian EHRs. The need for this study is best summed up by the House of Commons Report on the Health of the Lesbian, Gay, Bisexual, Trans, Queer, Intersex, Asexual and Two-Spirit (LGBTQIA2+) communities, which states that "...the LGBTQIA2+ communities in Canada experience numerous health inequities...data collection [should] be improved in order to obtain a more complete picture of the health of gender and sexual minorities in Canada." [21].

Methods

Definition of Terms

We conducted an environmental scan to examine public information sources on sex and gender definitions in Canadian EHRs and international health information standards. An environmental scan is a method of assessing the landscape on a specific topic based on multiple sources, including published literature, organizational documents, and key informants [22,23]. Sex and gender are broad terms used to include sex- and gender-related concepts such as sex at birth, gender identity,

and pronouns [5,24]. Definitions refer to data element names, code systems, and value sets for these concepts. The data element name is the label for an attribute, code system is the scheme used to codify concepts, and the value set is a group of codes an attribute can hold [25,26]. EHRs refer to electronic collections of an individual's lifetime health history and care records in the health ecosystem. They include input sources such as laboratory and pharmacy systems that collect individual health records, and administrative and clinical databases that house extracted records for health system analysis and reporting [27].

Information Sources

This study was built on an earlier environmental scan report from the Canada Health Infoway [28]. For that report, Canada Health Infoway staff extracted sex and gender definitions in EHR-related data dictionaries from public websites of Canadian health jurisdictions and national health care organizations. They also received guidance from the Infoway Sex and Gender Working Group on additional EHRs to be included. The Working Group is a multi-sectoral virtual community representing government agencies, health care organizations, advocacy groups, care providers, private vendors, and academics across Canada interested in sex and gender issues in EHRs. Drawing on the Infoway report, we added sex and gender definitions from international standards communities involved with the development, implementation, or approval of health information standards.

Data Analysis

For analysis, we summarized all sex and gender definitions from these sources and reviewed them for clarity, consistency, and accuracy. The review was conducted by the authors of this study with expertise or experience in patient care (KD), health information standards (FL), health equity (MA), and TGNB issues (RQ and KD). We also sought and received feedback on the definitions and earlier versions of this study from Infoway Sex and Gender Working Group members [12].

Results

Types of Information Sources

The information sources selected for this study consisted of public website descriptions of 52 databases and 55 data standards from 12 Canadian entities and 10 standards communities (Multimedia Appendix 1 [29-116]). The Canadian entities included (1) Alberta, British Columbia, Manitoba, Ontario, and Newfoundland and Labrador as 5 provincial jurisdictions responsible for 50 of these databases or data standards; (2) Canadian Institute for Health Information (CIHI) as the national health data holdings custodian responsible for 24 of these databases or standards; (3) Canadian Primary Care Sentinel Surveillance Network and Canadian Longitudinal Study

on Aging as 2 national consortiums responsible for the database for more than 450 primary care practices and the database for long-term community health survey of more than 50,000 aging individuals, respectively; (4) Statistics Canada (StatCan) as the federal agency responsible for the national data standards for population surveys, census, and cancer registries; (5) Canada Health Infoway as the national organization responsible for Canadian health data exchange and terminology standards; and (6) Centre for Addictions and Mental Health (CAMH) and Tri-Hospital and Toronto Public Health (Tri-H+TPH) as 2 health care organizations considered leaders and early adopters of sex and gender documentation in EHRs.

The standards communities included (1) Health Level Seven (HL7), Diagnostic Imaging and Communication in Medicine (DICOM), and the International Organization for Standardization as 3 international standards organizations that develop data exchange standards including HL7 Version 2 (V2) and Version 3 (V3), Clinical Document Architecture, Fast Healthcare Interoperable Resources (FHIR), and DICOM medical imaging standards; (2) Systemized Nomenclature of Medicine (SNOMED) International and Logical Observations Investigations Names and Codes (LOINC) as 2 international organizations that develop health terminology standards; (3) OpenEHR as an international collaboration that has developed the Gender Archetype, and the BioPortal where the Gender, Sex, and Sexual Orientation (GSSO) ontology developed by Kronk [111] is hosted; (4) ONC, National Health Services, and Australian Institute of Health and Welfare as 3 government agencies considered leaders in developing health information standards.

Sex and Gender Definitions in Existing Canadian EHRs

Sex

We extracted sex definitions and summarized them by data element, code system, and value set (Multimedia Appendix 2). There are 25 entries with 7 unique data element names. After merging names with minor variations, 5 distinct data element names remained. After eliminating duplicate entries, 7 code systems and 21 unique value sets remained (Textbox 1). Most variations are when sex is unknown, undifferentiated, other, or has no information. For instance, unknown is coded as *U*, *UNK*, *3*, or *39*; undifferentiated is coded as *UN*, *U*, or *1*. The same code can also have different values. For example, *U* can be *Unknown*, *Unknown or Undifferentiated*, *Undifferentiated*, or *Undetermined*; *UN* can be *Unknown*, *Undifferentiated*, or *Not assigned male or female*. Other can be *Other*, *Other (including hermaphrodites* and *transsexual)*, or *Other (person could not be uniquely identified)*. When no information is given, different codes and values are used, including *9-Not stated*, *281-Null*, and *blank*. From the definitions, it is unclear if *Undifferentiated*, *Undetermined*, and *Indeterminate* refer to the same concept.

Textbox 1. Summary of definitions for sex in Canadian electronic health records.

Data element name (n=5)

- Sex
- Administrative Sex
- Patient's sex
- Sex assigned at Birth
- Infant's sex

Code system (n=7)

- Health level Seven version 2 (HL7V2)-0001
- Diagnostic imaging and communication in medicine (0010,0040)
- Canadian Institute for Health Information
- Statistics Canada
- Newfoundland & Labrador Centre for Health Information
- Canadian Primary Care Sentinel Surveillance Network
- Manitoba

Value set (n=21)

- Male, female
- M-male, F-female
- M-male, F-female, blank (if unknown)
- M-male, F-female, I-indeterminate, UNK-unknown
- M-male, F-female, O-other
- M-male, F-female, O-other (trans-sexual, hermaphrodite)
- M-Male, F-Female, O-Other (person could not be uniquely identified as male or female; eg, hermaphrodite)
- M-male, F-female, U-unknown
- M-male, F-female, U-unknown, O-other
- M-male, F-female, U-unknown/undifferentiated, O-other
- M-male, F-female, U-undetermined, unknown or other
- M-male, F-female, U-undifferentiated, 9-not stated
- M-male, F-female, UN-not assigned male or female
- M-male, F-female, UN-undifferentiated, UNK-unknown
- 1-male, 2-female
- 1-male, 2-female, 3-other
- 1-male, 2-female, 3-unknown
- 1-male, 2-female, O-other
- 1-male, 2-female, O-other (includes hermaphrodites, transsexual)
- 1-male, 2-female, U-unknown
- 155,939-female, 133,338-male, 281-null, 39-unknown, 1-undifferentiated

---

Gender

We extracted gender definitions and summarized them by data element, code system, and value set (Multimedia Appendix 3).

There are 35 entries with 12 unique data element names. After merging names with minor variations, 5 distinct names remained. After eliminating duplicate entries, 9 code systems and 22 unique value sets remained (Textbox 2).

Textbox 2. Summary of definitions for gender in Canadian electronic health records.

Data element name (n=5)

- Gender

- Administrative Gender

- Clinical Gender

- Gender identity

- Newborn gender

Code system (n=9)

- Health Level Seven version 2

- Health Level Seven version 3

- Clinical Document Architecture

- Fast Healthcare Interoperable Resources

- Canadian Institute for Health Information

- Alberta

- Manitoba

- Newfoundland & Labrador

- Canadian Longitudinal Study on Aging

Value set (n=22)

- M-male, F-female

- M-male, F-female, D-gender diverse, UNK-not known, NA-not applicable

- M-male, F-female, I-undifferentiated stillbirth only, U-unknown, O-other; for trans-sexual or hermaphrodite, U-undifferentiated; for stillbirths only (discontinued)

- M-male, F-female, O-other (trans-sexual or hermaphrodites)

- M-male, F-female, OTH-other gender identity, UNK-not known, NA-not applicable

- M-male, F-female, U-unknown

- M-male, F-female, U-unknown, O-other

- M-male, F-female, U-unknown, O-other, I-indeterminate

- M-male, F-female, U-undifferentiated

- M-male, F-female, U-undifferentiated stillbirth only, O-other (trans-sexual or hermaphrodite)

- M-male, F-female, U-undifferentiated, stillbirths only, O-other or unknown

- M-male, F-female, UN-undifferentiated

- M-male, F-female, UN-undifferentiated, UNK-unknown

- M-male, F-female, UN-undifferentiated, UNK-unknown, NI-no information (could not be uniquely defined as male or female, eg, hermaphrodite), OTH-other

- M-male, F-female, 7-not collected, 9-unknown

- Male-male, Female-female, Unknown-unknown

- Male-male, Female-female, Unknown-unknown, Other-other

- Male-male, Female-female, refused or something else

- Male-male, Female-female, Other-other, Unknown-unknown

- 1-male, 2-female, 3-unknown

- 1-male, 2-female, 8-don't know/no answer, 9-refused

- 1-male, 2-female, ZZ-other

Most variations are in undifferentiated, unknown, other, and no information is given. For instance, undifferentiated is coded as *UN* or *U*; unknown as *Unknown*, *U*, *UNK*, *3*, or *9*; and other as *O*, *OTH*, *Other*, or *ZZ*. There are also different values for undifferentiated and other, regardless of whether the same code is used or not. For instance, the value for undifferentiated can be *Undifferentiated*, *Undifferentiated still birth only*, or *Undifferentiated (could not be uniquely defined as male or female, eg, hermaphrodite)*. Other can be *Other*, *Other or unknown*, *Other (trans-sexual or hermaphrodites)*, or *Other gender identity*. When no information is given, different codes

and values are used, including *7-Not collected*, *8-Don't know or no answer*, *9-Refused*, *Refused*, or *something else*, *NA-Not applicable*, or *NI-No information*. From the definitions, it is unclear if *Undifferentiated* and *Indeterminate* refer to the same concept.

Expanded Sex and Gender Definitions in Canadian Health Care Organizations

We extracted expanded sex and gender definitions in policy and practice guides of 4 Canadian institutions (Multimedia Appendix 4) and summarized them by organization, concept, and value set (Table 1).

Table 1. Expanded sex and gender definitions in Canadian health care organizations.

Concept (n=10)	Organization (n=4)	Value set and/or description if available (n=10)
Sex		
Sex	StatCan[a]	M-male, F-female
Variant of sex	StatCan	M-male, F-female, I-intersex
Sex at birth	CIHI[b]	M-male, F-female, I-indeterminate, UNK-unknown
Intersex	CIHI	Refers to a variety of conditions where a person has atypical development of sex characteristics, such as reproductive anatomy, sex chromosomes or sex-related hormones, that is not consistent with typical definitions of male of female
Gender		
	StatCan	M-male gender, F-female gender, D-gender diverse
	CIHI	M-male, F-female, D-gender diverse, UNK-unknown, NA-not applicable
	Tri_Hospital + TPH[c]	Male, female, trans-female to male, trans-male to female, intersex, other (specify), prefer not to answer, do not know
Cisgender		
	StatCan	C-cisgender, CM-cisgender man, CF-cisgender woman
Transgender		
	StatCan	T-transgender, TM-transgender man, TF-transgender woman, TG-transgender person, non-exclusive category
Nonbinary gender		
	StatCan	Transmasculine demi-boy, transfeminine demi-girl, pan-gender poly-gender bi-gender two-spirit, gender-fluid neutrois, genderless, agender
Lived gender		
	CIHI	How a person publicly presents their gender, which include behavior, appearances, name and pronoun
Gender identity		
	CIHI	A person's internal and experience, a sense of being a woman, man, both, neither or along a spectrum
	CAMH[d]	Male, female, transsexual, transgender, genderqueer, two-spirit, FTM (female-to-male), MTF (male-to-female), intersex, unsure, questioning, other, prefer not to answer

[a]StatCan: Statistics Canada.

[b]CIHI: Canadian Institute for Health Information.

[c]TPH: Toronto Public Health.

[d]CAMH: Centre for Addiction and Mental Health.

For sex, StatCan and CIHI have identified intersex as a distinct concept. For value sets, StatCan has defined *I-Intersex* as one of the sex options, whereas CIHI has *I-Indeterminate* as one of the sex at birth options. For gender, both StatCan and CIHI have *D-Gender Diverse* as a distinct concept. StatCan further distinguished gender into categories of cisgender, transgender,

and nonbinary gender with specific value sets such as *CM-Cisgender man*, *TM-Transgender man* and *Trans-Masculine*, and *Demi Boy*. CIHI has distinguished between gender identity and lived gender but only defined a value set for gender with *D-Gender Diverse* as an option. For CAMH and Tri-H+TPH, their gender value sets have many

options but with some variations: only CAMH has options for *Transsexual*, *Transgender*, *Genderqueer*, and *Questioning.* There are also variations when a person's sex or gender is not recorded: *UNK-Unknown*, *NA-Not Applicable*, *Other*, *Prefer not to answer*, or *Do not know.*

Sex and Gender Definitions in Standards Communities

Definitions

We extracted sex and gender definitions published by 10 standards communities (Multimedia Appendices 5 and 6). The published standards ranged from those in routine use such as HL7V2, V3, and DICOM specifications, to recently proposed specifications from the FHIR, HL7 Gender Harmony Project, ONC-Interoperability Standards Advisory, OpenEHR Gender Archetype, and BioPortal-GSSO. For sex, there are 11 unique data element names from 13 code systems and 16 value sets (Textbox 3). GSSO has also defined 3 sex subtypes as biological sex, anatomic sex, and genotypic sex. For gender, there are 10 unique data element names from 11 code systems and 12 value sets (Textbox 4).

Textbox 3. Summary of definitions for sex-related concepts from standards communities.

Data element names (n=11)

- Sex
- Administrative sex
- Patient's sex
- Sex of patients
- Patient sex (at birth)
- US-core-birthsex
- Sex assigned at birth
- Sex for clinical use
- Biological sex
- Anatomic sex
- Genotypic sex

Code systems (n=13)

- International Organization for Standardization-5218
- Australian Institute of Health and Welfare–Metadata Online Repository
- Diagnostic Imaging & Communication in Medicine (DICOM)–Context Identifier 7455
- DICOM- (0010,0040)
- DICOM- proposed
- BioPortal- gender sex and sexual orientation ontology
- Health Level 7 (HL7)V2-0001
- HL7V3
- Fast Healthcare Interoperability Resources
- National Health Services
- Logical Observations Investigations Names and Codes
- OpenEHR
- HL7 Gender Harmony

Value sets or subtypes (n=16)

- 1-male, 2-female, 0-not known, 9-not applicable
- 1-male, 2-female, 3-other, 9-not stated or inadequately described; alternate scheme
- M-male, F-female, X-other
- M-male, F-female, U-unknown sex, MP-male pseudohermaphrodite, FP-female pseudohermaprodite, H-hermaphrodite, MC-male changed to female, FC-female changed to male, 121104-ambiguous sex, 12102-other sex, 121103-undetermined sex
- M-male, F-female, A-ambiguous, N-not applicable, O-other, U-unknown
- M-male, F-female, O-other
- M-male, F-female, O-nonbinary, eg, intersex, other situations where neither male nor female apply clinically
- 1-male, 2-female, 8-not specified, 9-home leave (forward operational plans only)
- 76689-9 Sex assigned at birth
- M-male, F-female, UNK-unknown
- Male, female, intersex
- Male, female, nonbinary, unknown
- Sex subtypes (biological sex, anatomic sex, genotypic sex)
- Biological sex subtypes (chromosomal sex, anatomic sex, sex for imaging, hormonal or organ status sex)

- Anatomic sex subtypes (gonadal sex)

- Genotypic sex subtypes (chromosomal sex, genetic sex)

Textbox 4. Summary of definitions for gender-related concepts from standards communities.

Data element names (n=10)

- Gender

- Gender identity

- Administrative gender

- Person stated gender

- Code recorded gender or sex identity (previously legal gender)

- Legal gender

- Gender expression

- Affirmed gender

- Assigned gender

- Preferred pronoun

Code systems (n=11)

- Health Level 7 V3

- Fast Healthcare Interoperability Resources

- Diagnostic Imaging & Communication in Medicine

- Metadata Online Repository

- OpenEHR

- ISO-5218

- National Health Services

- Logical Observations Investigations Names and Codes

- Systemized Nomenclature of Medicine Clinical Terms

- Gender Sex and Sexual Orientation Ontology

- Office of National Coordinator Interoperability Standards Advisory

Value sets (n=12)

- M-male, F-female, UN-undifferentiated

- Male-male, Female-female, Other-other, Unknown-unknown

- 1-male, 2-female, 9-indeterminate (unable to be classified as either male or female), X-not known (not recorded)

- Man, woman, boy, girl, non-binary, notExpressed (examples only)

- 1-male (including trans man), 2-female (including trans woman), 3-non-binary, 4-other (not listed), Z-not stated (person asked but declined to provide a response)

- Transgender-female, transgender-male, non-binary, male, female, other, non-disclose

- 77691-5 gender identity

- 446151000124109|Male gender|, 446141000124107| Female gender|, 407377005|Femalte-to-male (FTM)/Transgender male/Trans man, 407376001|Male-to-female (MTF)/Transgender female/Trans woman|, 446131000124102| Identifies as non-confirming gender|, (US synonyms Include Genderqueer, Identifies as neither exclusively male nor female, Non-binary gender), OTH-Additional gender category or other, please specify, ASKU-Choose not to disclose

- Affirmed gender, assigned gender, assumed gender, authentic gender, felt gender, legal gender, natal gender

- Affirmed female, affirmed male

- Assigned female at birth, assigned male at birth

- Pronoun examples only—she, he, they, ze

Value Sets

There are variations in sex and gender value options. Examples are *M-Male* and *1-Male* for male, and *UN-Undifferentiated*, *9-Indeterminate*, *X-Other*, *X-Not known (not recorded)*, *0-Not known*, or *O-Non binary* if no information is available. When compared with more established standards such as HL7 V2 and DICOM, those from ONC, OpenEHR, and GSSO have more sex and gender value options. For instance, ONC has defined gender identity and sex assigned at birth value sets from LOINC and SNOMED CT. Examples include LOINC code *77691-5* for *Gender identity* and SNOMED CT code *407377005|Female-to-male transsexual|*. GSSO has more sex-related value sets based on biological features such as anatomical, chromosomal, brain, and genetic sex.

Discussion

Principal Findings

Overall Issues

There are wide variations in the definition and implementation of sex and gender in Canadian EHRs and international health information standards. There is a lack of clarity in some sex and gender concepts. There is inconsistency in the data element names, code systems, and value sets used to represent sex and gender concepts across EHRs. The appropriateness and adequacy of some value set options are questioned as our societal understanding of sexual health continues to evolve. Outdated value set options raise concerns about current EHRs supporting the provision of culturally competent, safe, and affirmative health care. The limited options available also perpetuate the inequities faced by the TGNB populations. The expanded sex and gender definitions from leading Canadian organizations and international standards communities have also brought implementation challenges in migrating these definitions into existing EHRs. These findings are elaborated below.

Lacking Conceptual Clarity

A major issue is the conflated use of sex and gender as a discrete single binary concept to define sex and gender identity. This longstanding practice can create confusion when a person's biology of being male or female along a spectrum is different from one's identity or feeling of one's gender identity as a woman, man, neither, both, or something else. The use of clinical gender, administrative gender, and administrative sex to represent sex and gender concepts for specific use is problematic because their context for use is often not well defined. As an example, the historical reliance on administrative gender as a means to assign hospital beds is inappropriate for transgender persons. The expanded sex and gender concepts from standard communities such as intersex, genotypic sex, and gender identity (eg, neutrois) are unfamiliar and sometimes overlapping terms that can be confusing to most but some clinicians and staff with specialty knowledge in sex and gender issues. These findings are consistent with the review by Jeffee [9], where care providers lacked familiarity with the terms and were ill-prepared to collect sex and gender data. They also echo the need for "accurate use of [sex- and gender-based medicine]

terminology in research and clinical practice across medical specialties and across health professions" [5].

Inconsistent Definitions

Similar to what is reported in the literature [3,117], there is a great deal of semantic variability in the data element names, code systems, and value sets used to represent sex and gender concepts in EHRs. For instance, sex, sex code, administrative sex, and patient's sex may all refer to a person's biology along a spectrum. However, different value set options exist across systems, including *indeterminate*, *intersex*, *undifferentiated*, and *not assigned male or female when the person is not male or female*. For gender, it is unclear if *undifferentiated*, *undifferentiated stillbirth only*, and *indeterminate* are equivalent terms. Even when sex and gender information is not available, it is unclear if such options as *unknown*, *no information*, *not collected*, *not applicable*, or *<blank>* have the same meaning. Although some value sets are from the same code systems such as DICOM and HL7, they can contain different options with little explanation (see Textbox 3). The HL7 Gender Harmony Project also noted these discrepancies in their recent work to standardize sex and gender definitions [118].

Inappropriate Value Options

As our understanding of sex and gender evolves over time, some historic value options are now considered offensive or outdated [119,120]. Terms such as *hermaphrodite* and *natal* are now recognized as being offensive. *M2F* and *F2M* are regarded as outdated by some and should be replaced by *transfeminine* and *transmasculine*. Similarly, *intersex* and *indeterminate* should not be interchangeable terms. Misuse of the label preferred pronouns over pronouns as a data element name is also common [9,120].

Evolving Value Options

As our society gives greater attention to sex and gender identity, the range of value options for sex and gender continues to expand. Dunne et al [7] reported that care providers and patients cited the need for a broader range of gender options and birth-assigned sex in EHRs. For use contexts, such terms as cisgender, transgender, and gender diversity may be adequate for classifying and reporting gender profiles at the population level. However, when caring for individual patients, distinguishing specific gender identity can be central to understanding health status, planning appropriate care, and maintaining patient safety. The proposed use of men and women as gender options instead of males and females also requires major shifts in society's discourse on sex and gender labeling [5,9]. As we broaden the range of value options available, the notion of third gender in other cultures such as Hijra in India also needs to be considered. To support culturally competent, equitable care requires expanding sex and gender options as an ongoing practice [7,121].

Implementation Challenges

There is no mandate in Canada on the type of sex and gender data to be collected in EHRs at present. Introducing expanded sex and gender definitions into existing EHRs can be a major undertaking for health care organizations, given the competing demands, limited resources, and redevelopment effort required.

Implementation efforts to date suggest that sex and gender concepts are foundational in nature, and implicate a myriad of technical, organizational, and societal aspects [17,122]. Technically, there is the potential for modernized definitions to affect all EHRs. As such, consensus on relevant sex and gender definitions is needed, and the implementation effort should be well coordinated. There needs to be a trusted source where one can validate the origin and provenance of the recorded sex and gender data in different EHRs to ensure their accuracy. Organizationally, the implementation plan and rationale for change must be clearly articulated to care providers, support staff, patients, and others affected. Education and training on the rationale, policies, and practices involved in collecting and using this information are necessary to cultivate a culturally safe and equitable environment and to prevent potential harm such as misgendering, outing, and deadnaming [5-7]. At the societal level, we need to engage the public to raise their awareness of this change and explain how it may affect their day-to-day interactions with the health ecosystem [9,15].

Proposed Actions

Need for Actions

On the basis of these findings, we propose a set of high-level actions to modernize sex and gender definitions in Canadian EHRs. These proposed actions are meant to initiate a dialogue with stakeholders, including individuals, communities, and organizations involved with developing or implementing, or affected by, the definitions. Collectively, they represent a set of guiding principles for organizations wishing to strive toward an inclusive EHR. These actions are also consistent with those reported by early adopters and leading organizations, including Fenway Health [17], Veteran Health Administration [11], University of California Davis Health System [122], and Canada's St. Michael's Hospital [123], that have started collecting expanded sex and gender data in their EHRs. Our proposed actions are described below.

Articulate the Need

Articulate the need for this work by explaining why the proposed actions are necessary to improve the health of sexual and gender minorities in Canada. These include the rationale for the collection of modernized sex and gender information in EHRs, how this information can advance our understanding of this population, and provide safe and culturally competent care. For instance, having expanded sex and gender definitions in the EHR is seen as an organization's obligation for inclusion as part of the Canadian Human Rights Code and Charter of Rights and Freedom [124].

Reach Consensus on Relevant Concepts

Reach consensus on relevant sex and gender concepts that can improve the health of Canadians, especially the underserved sex and gender minority populations. These concepts should address the need from policy, practice, research, and community perspectives [5]. Guidance from clinicians who use the data at the patient care level will be critical. This will help determine what codes and values are necessary versus what changes could occur at the policy level. Key concepts to consider include sex, sex assigned at birth, gender, administrative gender, gender

identity, pronouns, and such variants as current sex or gender, gender expression, and legal sex or gender.

Reach Consensus on Expanded Definitions in EHRs

Reach consensus on expanded sex and gender definitions in EHRs with respect to data element names, code systems, and value sets [14,26,27]. For instance, the use of subtypes and synonyms for data element names may improve shared meaning while allowing flexibility in local systems. Further harmonization of code systems and value sets to represent sex and gender concepts may facilitate wider adoption. An example is the development of reference sex-gender code systems and value sets with cross-maps to those used in local systems to guide the transition. Another is to allow write-in options at the patient care level that can be rolled-up for analysis and reporting without suppressing important data.

Develop a Coordinated Plan

Develop a coordinated action plan to implement expanded sex and gender definitions in Canadian EHRs. This plan should be action-oriented, stakeholder- and consensus-driven, and achievable in stages over time, given the limited resources and competing priorities in EHR redevelopment work [17,20]. The plan should also be coordinated with nonhealth government agencies where sex and gender data are collected (eg, driver's licenses and passports). Examples include expanding or harmonizing sex-gender concepts, creating reference sex-gender definitions with cross-maps to local systems, conducting small-scale implementation with selected EHRs, collaborating with standards communities, and sharing the lessons.

Embrace EHR Change From Socio-Organizational-Technical Aspects

Embrace EHR changes from socio-organizational and technical aspects to ensure success. Modernizing sex and gender definitions in EHRs is more than just a technical endeavor on data standardization but one that involves complex interplay of various social and organizational dynamics [125]. Organizations embarking on this journey need to engage senior executives, clinical leaders, and TGNB communities to cocreate an inclusive environment with culturally competent staff who are respectful of diversity, familiar with TGNB terminology, and knowledgeable in unique care needs of TGNB patients [126]. Key success factors include having explicit policy and practice guidelines on inclusion, ongoing staff education and training, a flexible EHR, and patient and public awareness to address the transition [11,15].

Demonstrate Benefits

Demonstrate the benefits in tangible terms by establishing relevant and meaningful measures to evaluate the impact of these expanded definitions in EHRs on individuals, communities, organizations, and the health ecosystem over time. The measures should be multifaceted covering policy, practice, research, and community perspectives. Knowledge translation is an essential element of this evaluation process to share the results and lessons learned. Examples of measures include adherence to best practices and respecting protected human rights, data completeness in EHRs, perceived usefulness by

patients and providers, and quantified improvements in health risks, service use, and health status.

Study Limitations

This study has some limitations. First, the information sources were restricted to public documents on existing EHRs and standards. To keep the study manageable, we excluded systems such as local EHRs, federal EHRs (eg, correctional facilities), and citizens' registries. Second, our study focused only on documented sex and gender definitions but not the process or context for their collection and use. The depth of our analysis is limited as there are deep societal, technical, ethical, and legal aspects that require more in-depth and thoughtful deliberations. Third, the environmental scan was based on the earlier Infoway report and key informant input. The work and results are not reproducible by others. Fourth, although we received valuable feedback from the virtual community, the study could benefit from wider consultations with other stakeholders concerned with sex and gender representation in EHRs. Finally, our proposed actions are presented only at a high level based on the findings of this scan. Sex and gender are complex, with implications for culture, technology, and society. Much work still lies ahead to design EHRs in ways that can support health equity, especially for the TGNB populations.

Conclusions

In this study, we examined sex and gender in Canadian EHRs and international health information standards. There are wide variations in how these concepts are represented and defined. The case to modernize sex and gender definitions in EHRs is compelling. Respecting the human rights aspect as well as safe, accurate, and efficient use, exchange, and reuse of sex and gender information for all populations is one condition for health equity that can be improved in a very direct way. A coordinated stakeholder- and consensus-driven action plan is needed for this effort to help advance the health of Canadians, especially the TGNB populations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of information sources.
[DOCX File , 52 KB - jmir_v22i11e20050_app1.docx]

Multimedia Appendix 2
Existing sex definitions in Canadian electronic health records.
[DOCX File , 30 KB - jmir_v22i11e20050_app2.docx]

Multimedia Appendix 3
Existing gender definitions in Canadian electronic health records.
[DOCX File , 31 KB - jmir_v22i11e20050_app3.docx]

Multimedia Appendix 4
Expanded sex and gender definitions in Canadian health care organizations.
[DOCX File , 30 KB - jmir_v22i11e20050_app4.docx]

Multimedia Appendix 5
Published definitions for sex-related concepts from standards communities.
[DOCX File , 31 KB - jmir_v22i11e20050_app5.docx]

Multimedia Appendix 6
Published definitions for gender-related concepts from standards communities.

References

1. Douglas MD, Dawes DE, Holden KB, Mack D. Missed policy opportunities to advance health equity by recording demographic data in electronic health records. Am J Public Health 2015 Jul;105(Suppl 3):S380-S388. [doi: 10.2105/AJPH.2014.302384] [Medline: 25905840]

2. Roberts TK, Fantz CR. Barriers to quality health care for the transgender population. Clin Biochem 2014 Jul;47(10-11):983-987. [doi: 10.1016/j.clinbiochem.2014.02.009] [Medline: 24560655]

3. Oemig F, Blobel B, Geibel P. How to work with vocabulary correctly, exemplified with gender coding? Stud Health Technol Inform 2016;228:344-348. [Medline: 27577401]

4. The Gender Harmony Project-Environmental Scan for Sex and Gender. InfoCentral - Canada Health Infoway. 2018. URL: https://infocentral.infoway-inforoute.ca/en/resources/docs/sex-gender/sex-gender-working-docs/3131-the-gender-harmony-project-environmental-scan-for-sex-and-gender [accessed 2020-10-20]

5. Madsen T, Bourjeily G, Hasnain M, Jenkins M, Morrison M, Sandberg K, et al. Article commentary: sex- and gender-based medicine: the need for precise terminology. Gender and the Genome 2017 Sep;1(3):122-128 [FREE Full text] [doi: 10.1089/gg.2017.0005]

6. Nadler LE, Ogden SN, Scheffey KL, Cronholm PF, Dichter ME. Provider practices and perspectives regarding collection and documentation of gender identity. J Homosex 2019 Sep 17 epub ahead of print. [doi: 10.1080/00918369.2019.1667162] [Medline: 31526306]

7. Dunne M, Raynor L, Cottrell E, Pinnock WJ. Interviews with patients and providers on transgender and gender nonconforming health data collection in the electronic health record. Transgend Health 2017;2(1):1-7 [FREE Full text] [doi: 10.1089/trgh.2016.0041] [Medline: 28861543]

8. Zhang Q, Goodman M, Adams N, Corneil T, Hashemi L, Kreukels B, et al. Epidemiological considerations in transgender health: a systematic review with focus on higher quality data. Int J Transgend Health 2020;21(2):125-137. [doi: 10.1080/26895269.2020.1753136] [Medline: 33015664]

9. Jaffee KD, Shires DA, Stroumsa D. Discrimination and delayed health care among transgender women and men: implications for improving medical education and health care delivery. Med Care 2016 Nov;54(11):1010-1016. [doi: 10.1097/MLR.0000000000000583] [Medline: 27314263]

10. Brown GR, Jones KT. Mental health and medical health disparities in 5135 transgender veterans receiving healthcare in the veterans health administration: a case-control study. LGBT Health 2016 Apr;3(2):122-131. [doi: 10.1089/lgbt.2015.0058] [Medline: 26674598]

11. Burgess C, Kauth MR, Klemt C, Shanawani H, Shipherd JC. Evolving sex and gender in electronic health records. Fed Pract 2019 Jun;36(6):271-277 [FREE Full text] [Medline: 31258320]

12. Goldstein Z, Corneil T, Greene DN. When gender identity doesn't equal sex recorded at birth: the role of the laboratory in providing effective healthcare to the transgender community. Clin Chem 2017 Aug;63(8):1342-1352. [doi: 10.1373/clinchem.2016.258780] [Medline: 28679645]

13. Gupta S, Imborek K, Krasowski MD. Challenges in transgender healthcare: the pathology perspective. Lab Med 2016 Aug;47(3):180-188 [FREE Full text] [doi: 10.1093/labmed/lmw020] [Medline: 27287942]

14. Deutsch M, Green J, Keatley J, Mayer G, Hastings J, Hall AM, World Professional Association for Transgender Health EMR Working Group. Electronic medical records and the transgender patient: recommendations from the world professional association for transgender health EMR working group. J Am Med Inform Assoc 2013;20(4):700-703 [FREE Full text] [doi: 10.1136/amiajnl-2012-001472] [Medline: 23631835]

15. Cahill S, Baker K, Deutsch MB, Keatley J, Makadon HJ. Inclusion of sexual orientation and gender identity in stage 3 meaningful use guidelines: a huge step forward for LGBT health. LGBT Health 2016 Apr;3(2):100-102. [doi: 10.1089/lgbt.2015.0136] [Medline: 26698386]

16. Woebkenberg J. Capturing sexual orientation/gender identity data through HIT. Journal Ahmia Org 2018 Nov 19:1-2 [FREE Full text]

17. Grasso C, McDowell M, Goldhammer H, Keuroghlian AS. Planning and implementing sexual orientation and gender identity data collection in electronic health records. J Am Med Inform Assoc 2019 Jan 1;26(1):66-70 [FREE Full text] [doi: 10.1093/jamia/ocy137] [Medline: 30445621]

18. Grasso C, Goldhammer H, Funk D, King D, Reisner SL, Mayer KH, et al. Required sexual orientation and gender identity reporting by us health centers: first-year data. Am J Public Health 2019 Aug;109(8):1111-1118. [doi: 10.2105/AJPH.2019.305130] [Medline: 31219717]

19. Deutsch MB, Keatley J, Sevelius J, Shade SB. Collection of gender identity data using electronic medical records: survey of current end-user practices. J Assoc Nurses AIDS Care 2014;25(6):657-663. [doi: 10.1016/j.jana.2014.04.001] [Medline: 24880490]

20. Deutsch MB, Buchholz D. Electronic health records and transgender patients: practical recommendations for the collection of gender identity data. J Gen Intern Med 2015 Jun;30(6):843-847 [FREE Full text] [doi: 10.1007/s11606-014-3148-7] [Medline: 25560316]

21. Casey B. The Health of LGBTQIA2 Communities in Canada. House of Commons of Canada. 2019. URL: https://www.ourcommons.ca/DocumentViewer/en/42-1/HESA/report-28/ [accessed 2019-06-01]

22. Graham P, Evitts T, Thomas-MacLean R. Environmental scans: how useful are they for primary care research? Can Fam Physician 2008 Jul;54(7):1022-1023. [Medline: 18625830]

23. Shahid M, Turin TC. Conducting comprehensive environmental scans in health research: a process for assessing the subject matter landscape. J Biomed Anal 2018 Sep 21;1(2):71-80 [FREE Full text] [doi: 10.30577/jba.2018.v1n2.13]

24. Rosendale N, Goldman S, Ortiz GM, Haber LA. Acute clinical care for transgender patients: a review. JAMA Intern Med 2018 Nov 1;178(11):1535-1543. [doi: 10.1001/jamainternmed.2018.4179] [Medline: 30178031]

25. Electronic Health Records. Canada Health Infoway. URL: https://www.infoway-inforoute.ca/en/solutions/digital-health-foundation/electronic-health-records [accessed 2020-10-20]

26. Gold S, Batch A, McClure R, Jiang G, Kharrazi H, Saripalle R, et al. Clinical concept value sets and interoperability in health data analytics. AMIA Annu Symp Proc 2018;2018:480-489 [FREE Full text] [Medline: 30815088]

27. HL7 Specification: Characteristics of a Value Set Definition, Release 1. Health Level 7. 2019. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=437 [accessed 2020-10-20]

28. Sex and Gender Working Group. Canada Health Infoway. URL: https://infocentral.infoway-inforoute.ca/en/wg/emha/collaboration/wg/sex-gender [accessed 2020-10-20]

29. Health Information Standards Committee - HISCA Standards. Alberta Government. URL: https://www.alberta.ca/hisca-standards.aspx [accessed 2020-10-26]

30. NACRS and CACS data standard. Version 1.0. Alberta Government. URL: https://open.alberta.ca/publications/nacrs-and-cacs-data-standard-version-1 [accessed 2020-10-26]

31. Chronic disease management data standard. Version 7.3. Alberta Government. URL: https://open.alberta.ca/publications/chronic-disease-management-data-standard-version-7-3 [accessed 2020-10-26]

32. Alberta Continuing Care Information System data standard. Version 1.0. Alberta Government. URL: https://open.alberta.ca/publications/alberta-continuing-care-information-system-data-standard-version-1-0 [accessed 2020-10-26]

33. Stakeholder client data standard. Version 1.4.2. Alberta Government. URL: https://open.alberta.ca/publications/stakeholder-client-data-standard-version-1-4-2 [accessed 2020-10-24]

34. Stakeholder demographic data standard. Version 2.2. Alberta Government. URL: https://open.alberta.ca/publications/stakeholder-demographic-data-standard-version-2-2 [accessed 2020-10-24]

35. Stakeholder provider data standard. Version 1.5. Alberta Government. URL: https://open.alberta.ca/publications/stakeholder-provider-data-standard-version-1-5 [accessed 2020-10-24]

36. Diagnostic imaging HL7 message specifications. Version 0.1. Alberta Government. URL: https://open.alberta.ca/publications/diagnostic-imaging-hl7-message-specifications-version-0-1 [accessed 2020-10-26]

37. Alberta diagnostic imaging reporting requirement data standard. Version 0.1. Alberta Government. URL: https://open.alberta.ca/publications/alberta-diagnostic-imaging-reporting-requirement-data-standard-version-0-1 [accessed 2020-10-26]

38. Medical summary for transfer of patient data. Version 0.1. Alberta Government. URL: https://open.alberta.ca/publications/medical-summary-for-transfer-of-patient-data-version-0-1 [accessed 2020-10-26]

39. Emergency health services : patient care reporting minimum data standard. Version 1.2. Alberta Government. URL: https://open.alberta.ca/publications/emergency-health-services-patient-care-reporting-minimum-data-standard [accessed 2020-10-26]

40. Adverse Events Following Immunization data standard [Version 2.2]. Alberta Government. URL: https://open.alberta.ca/publications/adverse-events-following-immunization-data-standard-version-2-2 [accessed 2020-10-24]

41. Breast cancer screening minimum data set. Version 0.8. Alberta Government. URL: https://open.alberta.ca/publications/breast-cancer-screening-minimum-data-set-version-0-8 [accessed 2020-10-24]

42. Communicable disease and outbreak data standard. Version 1.0. Alberta Government. URL: https://open.alberta.ca/publications/communicable-disease-and-outbreak-data-standard-v1-0 [accessed 2020-10-24]

43. Generic screening minimum data standard. Version 0.4. Alberta Government. URL: https://open.alberta.ca/publications/generic-screening-minimum-data-standard-version-0-4 [accessed 2020-10-24]

44. Alberta immunization data standard. Version 1.6. Alberta Government. URL: https://open.alberta.ca/publications/alberta-immunization-data-standard-version-1-6 [accessed 2020-10-26]

45. Newborn metabolic screening minimum data standard. Version 3.5. Alberta Government. URL: https://open.alberta.ca/publications/newborn-metabolic-screening-minimum-data-standard-version-3-5 [accessed 2020-10-26]

46. Perinatal health program data standard. Version 1.1. Alberta Government. URL: https://open.alberta.ca/publications/perinatal-health-program-data-standard-version-1-1 [accessed 2020-10-26]

47. Client Registry - Province of British Columbia. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/health-information-standards/standards-catalogue/client [accessed 2020-10-26]

48. British Columbia. Provider Registry. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/health-information-standards/standards-catalogue/client [accessed 2020-10-27]

49. British Columbia. Provider Laboratory Information Solution. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/health-information-standards/standards-catalogue/plis [accessed 2020-10-27]

XSL·FO
RenderX

50. British Columbia. Physician Information Technology Office. URL: https://www2.gov.bc.ca/assets/gov/health/practitioner-pro/pito-e2e-dtc-part2-consolidated-implementation-guide.pdf [accessed 2020-10-27]

51. British Columbia. Conformance Standards Version 3 Pharmanet. URL: https://www2.gov.bc.ca/assets/gov/health/practitioner-pro/software-development-guidelines/bc-conformance-standards-v3-pharmanet.zip [accessed 2020-10-27]

52. EMR Certification. Baseline EMR Requirements Specification. Shared Health Manitoba. 2019 Sep 13. URL: https://sharedhealthmb.ca/files/emr-certification-base-spec.pdf [accessed 2020-10-27]

53. EMR Certification. eChart Manitoba Context-Sensitive Launch Interface Specification. Shared Health Manitoba. 2019 Apr 31. URL: https://sharedhealthmb.ca/files/emr-certification-echart-spec.pdf [accessed 2020-10-26]

54. EMR Certification. eHealth_hub – Laboratory Result Distribution Interface Specification. Shared Health Manitoba. 2019 Apr 31. URL: https://sharedhealthmb.ca/files/emr-certification-hub-spec.pdf [accessed 2020-10-27]

55. EMR Certification. eHealth_hub – Diagnostic Imaging Report Result Distribution Interface Specification. Shared Health Manitoba. URL: https://sharedhealthmb.ca/files/emr-certification-hub-di-spec.pdf [accessed 2020-10-27]

56. EMR Certification eHealth_hub - Client Registry Query Service Interface Specification. Shared Health Manitoba. URL: https://sharedhealthmb.ca/files/emr-certification-hub-crq-spec.pdf [accessed 2020-10-27]

57. EMR Certification. Primary Care Quality Indicator Reminders and Data Extract Specification. Shared Health Manitoba. URL: https://sharedhealthmb.ca/files/emr-certification-pc-spec.pdf [accessed 2020-10-27]

58. EMR Certification. Interprofessional Team Data Extract Specification. Shared Health Manitoba. URL: https://sharedhealthmb.ca/files/emr-certification-de-spec.pdf [accessed 2020-10-27]

59. Provincial Discharge Abstract Database (PDADP) User Guide v.1.0. URL: https://www.nlchi.nl.ca/images/2016-17_PDAD_User_Guide_v1.0_2017-08-08.pdf [accessed 2020-10-27]

60. Live Birth System 2016 Master Methodology. URL: https://www.nlchi.nl.ca/images/2016_NLCHI_LBS_User_Guide_Final.pdf [accessed 2020-10-27]

61. Mortality System Data User Guide. URL: https://www.nlchi.nl.ca/images/2016_NLCHI_Mortality_User_Guide_Final.pdf [accessed 2020-10-27]

62. Stillbirth System 2016 User Guide. URL: https://www.nlchi.nl.ca/images/2016_NLCHI_Stillbirth_User_Guide_Final.pdf [accessed 2020-10-27]

63. Acute and Community CDR Input Standard Version. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/acute-and-community-cdr-input-standard [accessed 2020-10-27]

64. Ontario Oncology Information System Standards: Defining its Meaningful Use. Cancer Care Ontario. URL: https://www.cancercareontario.ca/sites/ccocancercare/files/guidelines/full/OncInfoSysStandards_0.pdf [accessed 2020-10-27]

65. Diagnostic Imaging Common Service Standard (DI CS). eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/diagnostic-imaging-common-service-standard-di-cs [accessed 2020-10-27]

66. Ontario Laboratories Information System (OLIS) Standard. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/ontario-laboratories-information-system-standard [accessed 2020-10-27]

67. Provincial Client Registry (PCR) Standards Release 2 (R2). eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/provincial-client-registry-standards-release-2-r2 [accessed 2020-10-27]

68. Provincial Clinical Document Architecture Header Standard. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/provincial-clinical-document-architecture-header-standard [accessed 2020-10-27]

69. Provincial Provider Registry Specification. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/provincial-provider-registry-specification [accessed 2020-10-26]

70. Digital Health Drug Repository Specification – FHIR. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/digital-health-drug-repository-specification-fhir [accessed 2020-10-27]

71. Digital Health Drug Repository Specification – FHIR (Release 3). eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/digital-health-drug-repository-specification-fhir-release-3 [accessed 2020-10-27]

72. Digital Health Immunization Repository Consumer Access Specification – FHIR. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/digital-health-immunization-repository-consumer-access-specification-fhir [accessed 2020-10-27]

73. Digital Health Immunization Repository Specification – FHIR (Release 1). eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/digital-health-immunization-repository-specification-fhir-release-1 [accessed 2020-10-27]

74. Digital Health Immunization Repository Specification – FHIR (Release 4). eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/digital-health-immunization-repository-specification-fhir-release-4 [accessed 2020-10-27]

75. Provincial Client Registry – HL7 FHIR Implementation Guide. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/provincial-client-registry-hl7-fhir-implementation-guide [accessed 2020-10-27]

76. Provincial Provider Registry. eHealth Ontario. URL: https://ehealthontario.on.ca/en/standards/provincial-provider-registry-fhir [accessed 2020-10-27]

77. EMR Specifications Library – EMR Core Data Set Version 5.0. OntarioMD. URL: https://www.ontariomd.ca/emr-certification/emr-specification/library [accessed 2020-10-27]

78. EMR Specifications Library – EMR Hospital Report Manager (HRM) Version 4.3. OntarioMD. URL: https://www.ontariomd.ca/emr-certification/emr-specification/library [accessed 2020-10-27]

79. Gender of people. Canada Health Infoway. URL: https://tgateway.infoway-inforoute.ca/singlesubset.html?id=2.16.840.1.113883.2.20.3.447 [accessed 2020-10-27]

80. Administrativegender. Canada Health Infoway. URL: https://tgateway.infoway-inforoute.ca/html/singlesubset.html?id=2.16.840.1.113883.2.20.3.308&versionid=20140930 [accessed 2020-10-27]

81. Data Holdings. Canadian Institute for Health Information. URL: https://www.cihi.ca/en/access-data-and-reports/make-a-data-request/data-holdings [accessed 2020-10-27]

82. CPCSSN Data Dictionary. Canadian Primary Care Sentinel Surveillance Network. URL: https://cpcssn.ca/wp-content/uploads/2015/02/CPCSSN-Data-Dictionary.pdf [accessed 2020-10-27]

83. Canadian Cancer Registry – Sex of person, category. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=assembleDESurv&DECId=467214&RepClass=591&Id=1211637&DFId=180443 [accessed 2020-10-27]

84. Classification of sex. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=469273 [accessed 2020-10-27]

85. Variant of classification of sex. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=469276 [accessed 2020-10-27]

86. Classification of gender. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=467245 [accessed 2020-10-27]

87. Classification of cisgender and transgender. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=469267&CVD=469268&CPV=C&CST=25012018&CLV=1&MLV=2 [accessed 2020-10-27]

88. Classification of cisgender and transgender. Statistics Canada. URL: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=469267&CVD=469268&CPV=T&CST=25012018&CLV=1&MLV=2 [accessed 2020-10-27]

89. Preparing Gender Identity for the National Statistical System. Statistics Canada. URL: http://www.asi-iea.ca/en/files/2018/09/Placemat_gender_ID_May12.pdf [accessed 2020-10-27]

90. Asking the Right Questions 2. Centre for Addictions and Mental Health (CAMH). URL: https://www.porticonetwork.ca/documents/489955/0/Asking+the+right+questions+2/80d19139-b361-4547-989d-e5b6479651b0 [accessed 2020-10-27]

91. We ask because we care, The Tri-Hospital + TPH Health Equity Data Collection Research Project Report– Summary Report. Tri-Hospital+Toronto Public Health. URL: https://www.porticonetwork.ca/documents/43843/277776/FINAL+Summary+Report+June+2013.pdf/2fc7d2ff-6573-427b-a247-9b347b1d55d9 [accessed 2020-10-27]

92. Gender of person who participant provided most care giving assistance. Canadian Longitudinal Study on Aging (CSLA). URL: https://datapreview.clsa-elcv.ca/mica/variable/com%3ACAG_GNDR_COM%3ACollected#/ [accessed 2020-10-27]

93. Data Definition Tables – Table 0001 Administrative Sex. Health Level 7. URL: https://www.hl7.org/special/committees/vocab/V26_Appendix_A.pdf [accessed 2020-10-27]

94. AdministrativeGender. Health Level 7. URL: http://www.hl7.org/v3ballotarchive/v3ballot/html/infrastructure/vocabulary/AdministrativeGender.html [accessed 2020-10-27]

95. Value Set AdministrativeGender. HL7 FHIR. URL: https://www.hl7.org/fhir/v3/AdministrativeGender/vs.html [accessed 2020-10-27]

96. U.S. Core Profiles Release 4.0.1. HL7 FHIR. URL: http://hl7.org/fhir/R4/valueset-administrative-gender.html [accessed 2020-10-27]

97. U.S. Core Profiles Release 3.1.1. us-core-birthsex. HL7 FHIR. URL: http://hl7.org/fhir/us/core/StructureDefinition-us-core-birthsex.html [accessed 2020-10-27]

98. UK Core Patient Profile – UK Core Person Stated Gender Code. HL7 FHIR. URL: https://simplifier.net/guide/ukcoredevelopment/allvaluesets#ValueSetUKCorePersonStatedGenderCode [accessed 2020-10-27]

99. Person Stated Gender Code. HL7 NHS Data Model and Dictionary. URL: https://datadictionary.nhs.uk/attributes/person_stated_gender_code.html [accessed 2020-10-27]

100. Gender Harmony Context Definitions. HL7 Gender Harmony Project. URL: https://confluence.hl7.org/display/VOC/Gender+Harmony+Context+Definitions [accessed 2020-10-27]

101. C.2.3 Patient Demographic Module - Patient's Sex. DICOM. URL: http://dicom.nema.org/dicom/2013/output/chtml/part03/sect_C.2.html [accessed 2020-10-27]

102. PS3.16.2020 Content Mapping Resources, CID 7455 Sex. DICOM. URL: http://dicom.nema.org/medical/dicom/current/output/html/part16.html#sect_CID_7455 [accessed 2020-10-27]

103. PS3.16 Correction Proposal, CP1927_14 Tag Patient's Sex and Patient's Gender. DICOM. URL: https://confluence.hl7.org/download/attachments/40743893/cp1927_14_Patient_sex_and_gender.docx?api=v2 [accessed 2020-10-27]

104. ISA 2020. Patient Gender Identity. ONC. URL: https://www.healthit.gov/isa/representing-patient-gender-identity [accessed 2020-10-27]

105. Interoperability Standards Advisory (ISA) 2020. Patient Sex (At Birth). Office of National Coordinator (ONC). URL: https://www.healthit.gov/isa/representing-patient-sex-birth [accessed 2020-10-27]

106. NHS Data Model and Dictionary – Gender Identity Code for Sexual Health. National Health Services (NHS). URL: https://datadictionary.nhs.uk/attributes/gender_identity_code_for_sexual_health.html [accessed 2020-10-27]

107. NHS Data Model and Dictionary – Gender Identity Same at Birth Indicator. National Health Services (NHS). URL: https://datadictionary.nhs.uk/attributes/gender_identity_same_at_birth_indicator.html [accessed 2020-10-27]

XSL•FO
RenderX

108. NHS Data Model and Dictionary – Sex of Patient. National Health Services (NHS). URL: https://datadictionary.nhs.uk/attributes/sex_of_patients.html [accessed 2020-10-27]

109. ISO/IEC 5218 Information Technology – Codes for the representation of human sexes. Wikipedia. URL: https://en.wikipedia.org/wiki/ISO/IEC_5218 [accessed 2020-10-27]

110. Gender Archetype. OpenEHR. URL: https://ckm.openehr.org/ckm/archetypes/1013.1.3715/mindmap [accessed 2020-10-27]

111. Kronk CA, Dexheimer JW. Development of the Gender, Sex, and Sexual Orientation ontology: Evaluation and workflow. J Am Med Inform Assoc 2020 Jul 01;27(7):1110-1115. [doi: 10.1093/jamia/ocaa061] [Medline: 32548638]

112. Metadata Online Registry (METeOR) - Gender. Australian Institute of Health and Welfare (AIHW). URL: https://meteor.aihw.gov.au/content/index.phtml/itemId/635994 [accessed 2020-10-27]

113. Metadata Online Registry (METeOR) - Sex. Australian Institute of Health and Welfare (AIHW). URL: https://meteor.aihw.gov.au/content/index.phtml/itemId/635126 [accessed 2020-10-27]

114. Sex and Gender. LOINC. URL: https://search.loinc.org/searchLOINC/userLogin.zul;jsessionid=815CC28182C507EF680491C98D28CD5A [accessed 2020-10-27]

115. Sex and Gender. SNOMED International. URL: https://browser.ihtsdotools.org/? [accessed 2020-10-27]

116. R4 Draft for Comment. Gender identity. HL7 FHIR. URL: https://hl7.org/fhir/2018Jan/codesystem-gender-identity.html [accessed 2020-10-27]

117. Moscoe G. Beyond the Binary: A Proposal for Uniform Standards for Gender Identity and More Descriptive Sex Classification in Electronic Medical Records. Digital Collections - OHSU. 2014 Dec. URL: https://digitalcollections.ohsu.edu/concern/etds/6d56zw83b?locale=en [accessed 2020-10-20]

118. McClure R. The Gender Harmony Project. Confluence Mobile - Confluence - HL7. 2020. URL: https://confluence.hl7.org/display/VOC/The+Gender+Harmony+Project [accessed 2020-10-20]

119. Victorian Family Violence Data Collection Framework: A Guide for the Collection of Family Violence Related Data by Victorian Government Departments, Agencies and Service Providers, Oct 2019. Victoria State Government. 2019 Oct. URL: https://www.vic.gov.au/about-victorian-family-violence-data-collection-framework [accessed 2020-10-20]

120. AHIMA Work Group. Improved patient engagement for LGBT populations: addressing factors related to sexual orientation/gender identity for effective health information management. J American Health Information Management Association 2017 Mar;88(3):34-39. [Medline: 29412543]

121. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: mixed-methods evaluation and recommendations. PLoS One 2017;12(5):e0178043 [FREE Full text] [doi: 10.1371/journal.pone.0178043] [Medline: 28542498]

122. Callahan E, Sitkin N, Ton H, Eidson-Ton S, Weckstein J, Latimore D. Introducing sexual orientation and gender identity into the electronic health record: one academic health center's experience. Acad Med 2015 Feb;90(2):154-160. [doi: 10.1097/ACM.0000000000000467] [Medline: 25162618]

123. Pinto AD, Aratangy T, Abramovich A, Devotta K, Nisenbaum R, Wang R, et al. Routine collection of sexual orientation and gender identity data: a mixed-methods study. Can Med Assoc J 2019 Jan 21;191(3):E63-E68. [doi: 10.1503/cmaj.180839] [Medline: 30665975]

124. Human Rights in Canada. Canadian Human Rights Commission. URL: https://www.chrc-ccdp.gc.ca/eng/content/human-rights-in-canada [accessed 2020-10-20]

125. Lau F, Antonio M, Davison K, Queen R, Devor A. A rapid review of gender, sex, and sexual orientation documentation in electronic health records. J Am Med Inform Assoc 2020 Sep 16 epub ahead of print. [doi: 10.1093/jamia/ocaa158] [Medline: 32935124]

126. Rosa W, Fullerton C, Keller R. Equality in Healthcare: The Formation and Ongoing Legacy of an LGBT Advisory Council. LGBT Health 2015 Dec;2(4):292-296. [doi: 10.1089/lgbt.2014.0068] [Medline: 26788769]

Abbreviations

CAMH: Centre for Addictions and Mental Health
CIHI: Canadian Institute for Health Information
CIHR: Canadian Institutes of Health Research
CT: clinical terms
DICOM: diagnostic imaging and communications in medicine
EHR: electronic health record
FHIR: fast healthcare interoperability resources
GSSO: gender, sex, and sexual orientation
HL7: health level seven
IT: information technology
LGBTQIA2+: lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual, two-spirit
LOINC: logical observation identifiers, names, and codes
ONC: Office of the National Coordinator

SNOMED: systemized nomenclature of medicine
StatCan: Statistics Canada
TGNB: transgender and nonbinary
TRI-H+TPH: Tri-Hospital and Toronto Public Health

XSL·FO

RenderX

Original Paper

Influence of Scanner Precision and Analysis Software in Quantifying Three-Dimensional Intraoral Changes: Two-Factor Factorial Experimental Design

Saoirse O'Toole[1], BA, BDentSc, PhD, MFD RCSI; David Bartlett[1], BDS (Hons), PhD, FDSRCS; Andrew Keeling[2], BSc (Hons), BDS, PhD; John McBride[3], BSc (Hons), PhD; Eduardo Bernabe[1], BDS, PhD; Luuk Crins[4], MSc, DDS; Bas Loomans[4], PhD, DDS

[1]Centre for Clinical, Oral and Translational Sciences, King's College London, London, United Kingdom

[2]University of Leeds, Leeds, United Kingdom

[3]University of Southampton, Southampton, United Kingdom

[4]Radboud University Medical Centre, Nijmegen, Netherlands

Corresponding Author:
Saoirse O'Toole, BA, BDentSc, PhD, MFD RCSI
Centre for Clinical, Oral and Translational Sciences
King's College London
London
United Kingdom
Phone: 44 2071887462
Email: saoirse.otoole@kcl.ac.uk

Abstract

Background: Three-dimensional scans are increasingly used to quantify biological topographical changes and clinical health outcomes. Traditionally, the use of 3D scans has been limited to specialized centers owing to the high cost of the scanning equipment and the necessity for complex analysis software. Technological advances have made cheaper, more accessible methods of data capture and analysis available in the field of dentistry, potentially facilitating a primary care system to quantify disease progression. However, this system has yet to be compared with previous high-precision methods in university hospital settings.

Objective: The aim of this study was to compare a dental primary care method of data capture (intraoral scanner) with a precision hospital-based method (laser profilometer) in addition to comparing open source and commercial software available for data analysis.

Methods: Longitudinal dental wear data from 30 patients were analyzed using a two-factor factorial experimental design. Bimaxillary intraoral digital scans (TrueDefinition, 3M, UK) and conventional silicone impressions, poured in type-4 dental stone, were made at both baseline and follow-up appointments (mean 36 months, SD 10.9). Stone models were scanned using precision laser profilometry (Taicaan, Southampton, UK). Three-dimensional changes in both forms of digital scans of the first molars (n=76) were quantitatively analyzed using the engineering software Geomagic Control (3D Systems, Germany) and freeware WearCompare (Leeds Digital Dentistry, UK). Volume change (mm^3) was the primary measurement outcome. The maximum point loss (μm) and the average profile loss (μm) were also recorded. Data were paired and skewed, and were therefore compared using Wilcoxon signed-rank tests with Bonferroni correction.

Results: The median (IQR) volume change for Geomagic using profilometry and using the intraoral scan was –0.37 mm^3 (–3.75-2.30) and +0.51 mm^3 (–2.17-4.26), respectively (P<.001). Using WearCompare, the median (IQR) volume change for profilometry and intraoral scanning was –1.21 mm^3 (–3.48-0.56) and –0.39 mm^3 (–3.96-2.76), respectively (P=.04). WearCompare detected significantly greater volume loss than Geomagic regardless of scanner type. No differences were observed between groups with respect to the maximum point loss or average profile loss.

Conclusions: As expected, the method of data capture, software used, and measurement metric all significantly influenced the measurement outcome. However, when appropriate analysis was used, the primary care system was able to quantify the degree of change and can be recommended depending on the accuracy needed to diagnose a condition. Lower-resolution scanners may underestimate complex changes when measuring at the micron level.

Introduction

All clinicians should be able to quantify and assess whether a degenerative health condition is stable or progressing. This is possible in some diseases that have accurate biomarkers but is not always possible for diseases of the soft and hard tissues. Measurement has typically taken the form of recording subjective visual changes; therefore, physical measurements of change are needed.

In dentistry, quantitative measurement of differences between sequential 3D scans of the teeth is typically used to diagnose erosive tooth wear. This is a condition in which excessive acids from the diet and stomach can dissolve the teeth. Due to changes in diet and health, the prevalence of erosive tooth wear has been increasing, now affecting 1 in every 3 adults globally [1]. Quantitative validation of tooth wear has only been possible in university hospitals thus far [2-5]. This validation is achieved by scanning accurate molds of the teeth with laser profilometers to create a precise digital map of the surface with repeatable, calibrated point coordinates. As directly scanning teeth with lab-based profilometers has not been possible, scans of molds of the teeth have been aligned and compared using custom-built or commercial engineering software to quantify changes. Data capture is typically very accurate with this approach, and small process errors have been calculated to be in the range of 15 microns [6,7]. However, the reliance on research laboratory–based scanners and complex engineering analysis software is expensive and unfeasible for use in primary care settings [6-8].

Digital handheld scanners, known as intraoral scanners, take digital maps of the teeth and are increasingly being used in primary care. Intraoral scanners do not generate aerosols and are more amenable to effective cross-infection control compared to conventional impressions that generate aerosols and can harbor pathogenic microorganisms [9]. These advantages are particularly relevant for ensuring complete disinfection during the current COVID-19 pandemic. Intraoral scanners capture data via different methods, ranging from video capture to the use of confocal, triangulation, or active wavefront principles. Rather than relying on accurate calibrated data point collection on an unmoving subject, multiple data points are captured and stitched together with company-specific algorithms. Errors are generated when the scanner fails to collect sufficient data to stitch a digital map of the surface (undersampling) [10,11] or when the process fails, particularly with more than one tooth [12,13]. Furthermore, data stitching algorithms often interpolate or smooth missing or erroneous data; therefore, the data points are estimated, nonuniform, and lack adequate surface detail for changes to be measured at the micron level.

The software currently used by commercial companies to analyze digital maps rely on an iterative closest point (ICP) algorithm to merge the maps to the closest possible alignment,

without considering if the proposed alignment solution makes biological sense [14]. We previously demonstrated that this estimation leads to distortions and can result in physiologically impossible outcomes [15,16]. We recently incorporated feature-recognizing elements [16,17] into an ICP algorithm to minimize these errors and created an open-source freeware to be used alongside any 3D scan. Although this method has been validated against previous gold-standard software [18], it has not yet been tested on longitudinal clinical data.

The combination of data collection from primary care and free, user-friendly software for analysis may create new opportunities for monitoring disease. However, the accuracy of measuring change in scans will be influenced by the scanner, software, and their interaction. In this study, we used a factorial design to compare data obtained from profilometric scans of casts and those obtained from direct intraoral scans using two types of registered software: commercial software (Geomagic Control, 3D Systems, Germany) and freeware (WearCompare, Leeds Digital Dentistry, UK). We expected to see differences in the measurements obtained between the scanners but we did not know whether this difference would be clinically significant. The primary null hypothesis was that the dental wear data, specifically the volume change, average profile loss, and maximum point loss, detected by the profilometer will not be different to those obtained with the intraoral scanner. The secondary null hypothesis was that the software used to analyze the data will not influence the volume change, average profile loss, and maximum point loss observed for either scanner.

Methods

Participants

Data were collected from a larger clinical longitudinal erosive tooth wear study (Radboud Tooth Wear Project ABR code: NL31371.091.10) [19,20]. Study participants had been referred by general dental practitioners to the Department of Dentistry of Radboud University Medical Center (Nijmegen, the Netherlands) for management of erosive tooth wear. Patients in the monitoring arm who provided additional written consent for their data to be transported to the United Kingdom (ABR codes NL31401.091.10) and to perform additional analyses were included in the study (n=25; mean age 35.8, SD 6.8 years; 20 men and 5 women). A power calculation was performed in GPower vs 3.1 [21] using a two tailed test, demonstrating that to obtain a correlation of 0.4 between the scanners at 95% power with $P<.05$, a sample size of 75 was required.

Digital and Dental Impressions

The data transported to the United Kingdom included digital intraoral scans obtained using Lava Chairside Oral Scanner (3M, USA) at baseline and 3M True Definition Intraoral Scanner (3M, UK) at follow up, and analog dental impressions taken with silicone (Ivoclar Virtual 380, Ivoclar Vivodent, Liechtenstein, Europe). Impressions were poured in type-3

XSL•FO

RenderX

dental stone (SLR Dental GmbH, Germany) within 24 hours according to the manufacturer's instructions.

Both the digital and dental impressions were captured by the same trained operator. The point clouds of recognized index teeth (ie, the occlusal surface of the first molars [22,23]) were isolated by the operator (ST) and set aside for evaluation. Each analog study model (n=100) was scanned using a noncontact triangulation laser profilometer (XYRIS 2000TL, Taicaan Technologies, Southampton, UK) in a raster pattern using a step-over of 50 μm with a repeatability error of 2.6 μm [24]. This generated a 3D point cloud dataset for comparison.

Measurements and Software

Quantitative analysis of the change between sequential scans from the profilometer and intraoral scanner was performed using both the commercial software Geomagic Control 2011 (Geomagic, Morrisville, NC, USA) and the freeware WearCompare (Leeds Digital Dentistry, Leeds, UK). Data points selected by the operator on the buccal and lingual surfaces were chosen as reference areas and used for analysis according to previously published protocols [16]. For Geomagic, a best-fit alignment of 1000 data points on reference surfaces, followed by a refined alignment using 5000 data points, was performed. For the reference alignment, the occlusal surface was deleted from the dataset, leaving the buccal and lingual reference surfaces. The transformation matrix was then applied to the complete displaced dataset to realign it with the same orientation. For WearCompare, an initial global alignment utilizing a feature-based recognition system was performed. The same buccal and lingual reference surfaces were selected for refined ICP alignments, which highlights corresponding reference areas within 25 microns of each other. The occlusal surface was selected to be measured and all measurements were taken perpendicular to the occlusal surface.

Volume change (mm^3), maximum point loss (μm), and the mean loss over the surface (μm) were analyzed for each surface for both scanners and software types. As a secondary volumetric analysis, any positive values, indicating either gain or error, were set to zero.

Statistical Analysis

This study utilized a two-factor factorial experimental design comparing two different methods of data capture (profilometer and intraoral scanner) and two different analysis software types with different alignment principles (Geomagic Control and WearCompare). Descriptive statistics of all measurement metrics were calculated, and normality was assessed using the Shapiro-Wilks test and histogram assessment. Since the data were paired and skewed, Wilcoxon signed-rank tests were used

to compare outcomes (volume change, maximum point loss, and the average loss over the surface) between groups. Bonferroni correction was applied to compensate for multiple comparisons. The significance level was set at .008 (.05/6) to identify differences between groups. Single-measures intraclass correlation (ICC) analysis was performed between data capture method (scanner) and data analysis method (software). All analyses were performed in SPSS version 25 (IBM Corporation, Armonk, NY, USA).

Results

From the original data collected in the Netherlands, 76 surfaces were analyzed representing an average follow-up time of 36 months (SD 10.9). The data were initially analyzed using the previous gold-standard commercial software Geomagic. Laboratory profilometry data analyzed in Geomagic showed a median volume loss of –0.37 mm^3 (IQR –3.75-2.30), whereas a median volume gain was observed for the intraoral scan data of +0.51 mm^3 (IQR –2.17-4.26), representing a significant difference (P<.001). The median profile loss was 55.8 μm (IQR 24.43-77.60) and 43.65 μm (IQR 29.93-77.95) for the laboratory profilometer and intraoral scanner, respectively (P=.001). The maximum point loss on the occlusal surface was 398.4 μm (IQR 238.7-533.7) and 303.9 μm (IQR 217.6-483.0) for the profilometer and intraoral scan, respectively (P=.01).

Data from the freeware WearCompare showed a median volume change for the profilometer scan of –1.21 mm^3 (IQR –3.48-0.56) and –0.39 mm^3 (IQR –3.96-2.76) for the intraoral scan (P=.04). The median profile loss was 44.80 μm (IQR 29.48-91.63) and 43.10 μm (IQR 24.43-77.60) for the profilometer scan and intraoral scan, respectively (P=.18). The median maximum point loss on the occlusal surface for the profilometer scan was 317.1 μm (IQR 198.0-466.4) and was 278.3 μm (IQR 170.8-494.0) for the intraoral scan (P=.77). Therefore, no statistically significant differences were observed between the profilometer scans and intraoral scans when measurements were analyzed in WearCompare.

However, WearCompare detected significantly greater volume loss than analysis in Geomagic (P<.001), regardless of the scanner type. There were no differences between software in terms of average profile loss (P=.28) or maximum point loss (P=.26). When positive values were set to zero, the median volume change for the profilometer was unchanged, whereas intraoral scan volume loss evident. A significant difference was observed between the profilometry and intraoral scan data for Geomagic analysis (P=.02) but not for WearCompare analysis (P=.36) (Table 1).

Table 1. Volume changes (mm^3) observed over 3 years when positive data were set to zero, which is a commonly used method by many commercial companies.

Scan type	Geomagic	WearCompare	P value
Laser profilometer, median (IQR)	–0.37 (–3.75-0.00)	–1.21 (–3.48-0.00)	.30
Intraoral scanner, median (IQR)	0.00 (–2.17-0.00)	–0.39 (–3.95-0.00)	.09

XSL•FO
RenderX

Moderate ICCs were observed in analyzing volume change data between the scanners and software types: 0.476 (95% CI 0.281-0.632) for Geomagic and 0.457 (95% CI 0.259-0.618) for WearCompare (both $P<.001$). WearCompare and Geomagic data showed slightly stronger ICC using the intraoral scanner (0.673, 95% CI 0.529-0.780; $P<.001$) than with the profilometer (0.525, 95% CI 0.341-0.671, $P<.001$).

Discussion

Principal Findings

This study demonstrates the differences in outcomes that can be observed when using low-resolution primary care digital scanners and precision measurements from hospital laboratory profilometers for measuring biological changes at the micron level. As expected, increased volume change values were observed using the higher-resolution and calibrated profilometer scans compared to the intraoral scans. Surprisingly, this difference was only statistically significant when using commercial software, previously considered to be the gold standard, for the analysis. The custom-built freeware outperformed the commercial software. The null hypothesis was therefore partially rejected. This finding suggests that if the analysis is conducted accurately, it may compensate for the decreased resolution of the scanner. This is a promising finding and has implications for the development of primary care systems.

There are several possible reasons for the reduced volume changes observed with intraoral scanners. Data interpolation or the mathematical averaging of data points across a surface can smooth the topography of the surface and may overlook small discrepancies/areas of change in the surface. Smooth surface lesions, potentially on the buccal and lingual reference areas, will be subjected to heavy data undersampling internally in the scanner as the topography is not deemed to be as important. Smoothened surfaces are more susceptible to inaccuracies in data registration and alignment [15] as they will increase the mathematical tendency to minimize differences toward any sloped surfaces (in this case the occlusal surface). This can result in inaccuracies in alignment and biologically implausible outcomes. Analysis in a software that ignores features or the holistic geometric shape, such as Geomagic used in this study, will be particularly susceptible to this effect. Combining Geomagic analysis with the intraoral scan data resulted in an overall volumetric tooth tissue gain, which is physiologically impossible, indicating large errors within the analysis process.

For WearCompare, errors did not occur to the same extent, resulting in overall negative values for wear progression in both the profilometry and intraoral scans. Recent techniques developed at Radboud University involve using reference areas for alignment on the occlusal surfaces in addition to the buccal and lingual surfaces, providing additional control of the alignment along the Z axis. This may facilitate less translation and angular errors, and consequently less positive values. However, this comes at a tradeoff of increased analysis time and may also underestimate wear if an ICP algorithm is used for the Z axis. Further research will focus on validating this technique.

The correlation between wear measurements taken with the scanners was moderate as there are inherent but different errors for each form of data capture. Undetected, subvisual errors on casts or scans may have been present and subsequently analyzed as wear data. The profilometer is unable to scan undercuts, which means that less of the surface area can be used for selective surface alignment. In contrast, the intraoral scanner was successful at scanning undercuts. However, missing data or incomplete intraoral scans can also create errors whereby the triangle size is distorted and measurements can be skewed [25]. Recognizing where such errors may lie in each scan type will facilitate more accurate analysis.

Differences were observed between the profilometer and intraoral scan data when positive values were omitted from the analysis. Discounting positive values, which is commonly done with many types of commercial software and when profile loss and maximum point loss measurement metrics are reported, do not show error within the system. We observed that this can cause clinically significant changes to the outcome. Color maps of aligned scans can visually indicate areas of change, but the quantification does not always reflect the severity of progression. Discounting positive data increases the likelihood that a poor alignment will not be detected and wear underestimation or overestimation can occur. Reporting the negative changes may only be useful when trying to communicate wear to patients, but these metrics have limited diagnostic potential when measuring successive rates of wear.

Limitations

This study has several limitations. The analysis was performed blinded to the sequence of scanning to limit bias. However, there were often indications of sequence, such as surface restorations or clear visual wear progression in the interim period. Although only one model of intraoral scanner was used, the hardware and software changed over the 3-year period of the study, emphasizing that research in this fast-moving field becomes rapidly outdated. Other intraoral scanners will have slightly different methods of processing missing data and interpolating irregularities, and it is possible that slightly different results may be achieved with different intraoral scanners. Single-tooth analysis was performed to maximize accuracy, which also limits generalizability to full-arch analysis. A large limitation in longitudinal wear analysis is that the true wear progression is unknown. One has to assume that wear has occurred and positive values represent errors in alignment or in the data capture process. This makes it difficult to identify any form of measurement as a gold standard.

Conclusion

This study shows that low-resolution scanners can be used for measurements at the micron level provided appropriate analysis techniques and software are used. This could represent a step change in the way that erosive tooth wear is diagnosed and treated. From a dental point of view, the ability to view digital scans with increased magnification on a monitor also offers an increased diagnostic advantage. However, there is a duty of care on the profession and research community to not overestimate the quantitative capabilities of digital scanners to inform treatment or care outcomes until we are certain that they are

adequately sensitive and specific to do so. This will depend on the level of accuracy required from the analysis process to diagnose disease progression within a feasible diagnostic window. The resolution and accuracy of primary care scanning tools is likely to increase rapidly and further work should concentrate on reducing the process errors inherent within each measurement system.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Schlueter N, Luka B. Erosive tooth wear - a review on global prevalence and on its prevalence in risk groups. Br Dent J 2018 Mar 09;224(5):364-370. [doi: 10.1038/sj.bdj.2018.167] [Medline: 29495027]
2. Rodriguez J, Austin R, Bartlett D. In vivo measurements of tooth wear over 12 months. Caries Res 2012;46(1):9-15. [doi: 10.1159/000334786] [Medline: 22156738]
3. Mitchell HL, Chadwick RG, Ward S, Manton SL. Assessment of a procedure for detecting minute levels of tooth erosion. Med Biol Eng Comput 2003 Jul;41(4):464-469. [doi: 10.1007/bf02348090]
4. Pintado MR, Anderson GC, DeLong R, Douglas WH. Variation in tooth wear in young adults over a two-year period. J Prosthet Dent 1997 Mar;77(3):313-320. [doi: 10.1016/s0022-3913(97)70189-6]
5. O'Toole S, Newton T, Moazzez R, Hasan A, Bartlett D. Randomised Controlled Clinical Trial Investigating The Impact of Implementation Planning on Behaviour Related to The Diet. Sci Rep 2018 May 23;8(1):8024-8026. [doi: 10.1038/s41598-018-26418-0] [Medline: 29795123]
6. Rodriguez JM, Bartlett DW. The dimensional stability of impression materials and its effect on in vitro tooth wear studies. Dent Mater 2011 Mar;27(3):253-258. [doi: 10.1016/j.dental.2010.10.010] [Medline: 21075442]
7. Chadwick RG, Mitchell HL, Ward S. Evaluation of the accuracy and reproducibility of a replication technique for the manufacture of electroconductive replicas for use in quantitative clinical dental wear studies. J Oral Rehabil 2002 Jun;29(6):540-545. [doi: 10.1046/j.1365-2842.2002.00894.x] [Medline: 12071922]
8. Rodriguez J, Austin R, Bartlett D. A method to evaluate profilometric tooth wear measurements. Dent Mater 2012 Mar;28(3):245-251. [doi: 10.1016/j.dental.2011.10.002] [Medline: 22094292]
9. Barenghi L, Barenghi A, Cadeo C, Di Blasio A. Innovation by Computer-Aided Design/Computer-Aided Manufacturing Technology: A Look at Infection Prevention in Dental Settings. Biomed Res Int 2019;2019:6092018. [doi: 10.1155/2019/6092018] [Medline: 31467901]
10. Shembesh M, Ali A, Finkelman M, Weber H, Zandparsa R. An In Vitro Comparison of the Marginal Adaptation Accuracy of CAD/CAM Restorations Using Different Impression Systems. J Prosthodont 2017 Oct 08;26(7):581-586. [doi: 10.1111/jopr.12446] [Medline: 26855068]
11. Keeling A, Wu J, Ferrari M. Confounding factors affecting the marginal quality of an intra-oral scan. J Dent 2017 Apr;59:33-40. [doi: 10.1016/j.jdent.2017.02.003] [Medline: 28189718]
12. Güth JF, Edelhoff D, Schweiger J, Keul C. A new method for the evaluation of the accuracy of full-arch digital impressions in vitro. Clin Oral Investig 2016 Sep 10;20(7):1487-1494. [doi: 10.1007/s00784-015-1626-x] [Medline: 26454734]
13. Giménez B, Özcan M, Martínez-Rus F, Pradíes G. Accuracy of a digital impression system based on parallel confocal laser technology for implants with consideration of operator experience and implant angulation and depth. Int J Oral Maxillofac Implants 2014 Jul;29(4):853-862. [doi: 10.11607/jomi.3343] [Medline: 25032765]
14. Mitchell HL, Chadwick RG. Mathematical shape matching as a tool in tooth wear assessment--development and conduct. J Oral Rehabil 1998 Dec 22;25(12):921-928. [doi: 10.1046/j.1365-2842.1998.00324.x] [Medline: 9888227]
15. Mitchell HL, Koch I, Chadwick RG. Linear interpolation error in measured surfaces in a dental erosion study. Med Biol Eng Comput 2004 Jan;42(1):100-105. [doi: 10.1007/bf02351017]
16. O'Toole S, Osnes C, Bartlett D, Keeling A. Investigation into the accuracy and measurement methods of sequential 3D dental scan alignment. Dent Mater 2019 Mar;35(3):495-500. [doi: 10.1016/j.dental.2019.01.012] [Medline: 30683418]
17. Zhou QY, Park J, Koltun V. Fast global registration. Cham: Springer; 2016 Presented at: 14th European Conference on Computer Vision; October 2016; Amsterdam, Netherlands p. 766-782. [doi: 10.1007/978-3-319-46475-6_47]
18. O'Toole S, Osnes C, Bartlett D, Keeling A. Investigation into the validity of WearCompare, a purpose-built software to quantify erosive tooth wear progression. Dent Mater 2019 Oct;35(10):1408-1414. [doi: 10.1016/j.dental.2019.07.023] [Medline: 31402133]
19. Sterenborg BAMM, Bronkhorst EM, Wetselaar P, Lobbezoo F, Loomans BAC, Huysmans MDNJM. The influence of management of tooth wear on oral health-related quality of life. Clin Oral Investig 2018 Sep 3;22(7):2567-2573 [FREE Full text] [doi: 10.1007/s00784-018-2355-8] [Medline: 29397468]

20.   Loomans B, Opdam N. A guide to managing tooth wear: the Radboud philosophy. Br Dent J 2018 Mar 09;224(5):348-356. [doi: 10.1038/sj.bdj.2018.164] [Medline: 29495023]

21.   Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

22.   Martignon S, López-Macías AM, Bartlett D, Pitts N, Usuga-Vacca M, Gamboa LF, et al. The use of index teeth vs. full mouth in erosive tooth wear to assess risk factors in the diet: A cross-sectional epidemiological study. J Dent 2019 Sep;88:103164. [doi: 10.1016/j.jdent.2019.07.002] [Medline: 31276748]

23.   O'Toole S, Lau JS, Rees M, Warburton F, Loomans B, Bartlett D. Quantitative tooth wear analysis of index teeth compared to complete dentition. J Dent 2020 Jun;97:103342. [doi: 10.1016/j.jdent.2020.103342] [Medline: 32360512]

24.   Rodriguez JM, Bartlett DW. A comparison of two-dimensional and three-dimensional measurements of wear in a laboratory investigation. Dent Mater 2010 Oct;26(10):e221-e225. [doi: 10.1016/j.dental.2010.07.001] [Medline: 20701961]

25.   Nedelcu R, Olsson P, Nyström I, Thor A. Finish line distinctness and accuracy in 7 intraoral scanners versus conventional impression: an in vitro descriptive comparison. BMC Oral Health 2018 Feb 23;18(1):27-11 [FREE Full text] [doi: 10.1186/s12903-018-0489-3] [Medline: 29471825]

Abbreviations

ICC: intraclass correlation
ICP: iterative closest point

XSL•FO

RenderX

Original Paper

Advanced Symptom Management System for Patients with Malignant Pleural Mesothelioma (ASyMSmeso): Mixed Methods Study

Roma Maguire[1], BN, MSc, PhD; John Connaghan[1], MA, MSc; Anne Arber[2], MSc, PhD; Naomi Klepacz[2], BSc, MSc, PhD; Kevin G Blyth[3,4], MBChB, MD; John McPhelim[5], BSc; Paul Murray[6], MA, MD; Hitasha Rupani[7], BM, MRCP, PhD; Anoop Chauhan[7], MB, ChB, PhD; Peter Williams[8], BA, MSc; Laura McNaughton[3], BSc; Kirstie Woods[9], MSc; Anne Moylan[7], BSc

[1]Department of Computing and Information Sciences, University of Strathclyde, Glasgow, United Kingdom

[2]Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[3]Glasgow Pleural Disease Unit, Queen Elizabeth University Hospital, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom

[4]Institution of Cancer Sciences, University of Glasgow, Glasgow, United Kingdom

[5]University Hospital Hairmyres, NHS Lanarkshire, East Kilbride, United Kingdom

[6]Ashford and St Peter's NHS Trust, Surrey, United Kingdom

[7]Portsmouth Hospitals University NHS Trust, Portsmouth, United Kingdom

[8]Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, United Kingdom

[9]University Hospital Monklands, NHS Lanarkshire, Airdrie, United Kingdom

Corresponding Author:
Roma Maguire, BN, MSc, PhD
Department of Computing and Information Sciences
University of Strathclyde
LT14 14a, Livingstone Tower
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 (0)141 548 3589
Email: Roma.maguire@strath.ac.uk

Abstract

Background: Patients with malignant pleural mesothelioma (MPM) have a life-limiting illness and short prognosis and experience many debilitating symptoms from early in the illness. Innovations such as remote symptom monitoring are needed to enable patients to maintain wellbeing and manage symptoms in a proactive and timely manner. The Advanced Symptom Management System (ASyMS) has been successfully used to monitor symptoms associated with cancer.

Objective: This study aimed to determine the feasibility and acceptability of using an ASyMS adapted for use by patients with MPM, called ASyMSmeso, enabling the remote monitoring of symptoms using a smartphone.

Methods: This was a convergent mixed methods study using patient-reported outcome measures (PROMs) at key time points over a period of 2-3 months with 18 patients. The Sheffield Profile for Assessment and Referral for Care (SPARC), Technology Acceptance Model (TAM) measure for eHealth, and Lung Cancer Symptom Scale-Mesothelioma (LCSS-Meso) were the PROMs used in the study. Patients were also asked to complete a daily symptom questionnaire on a smartphone throughout the study. At the end of the study, semistructured interviews with 11 health professionals, 8 patients, and 3 carers were conducted to collect their experience with using ASyMSmeso.

Results: Eighteen patients with MPM agreed to participate in the study (33.3% response rate). The completion rates of study PROMs were high (97.2%-100%), and completion rates of the daily symptom questionnaire were also high, at 88.5%. There were no significant changes in quality of life, as measured by LCSS-Meso. There were statistically significant improvements in the SPARC psychological need domain (P=.049) and in the "Usefulness" domain of the TAM (P=.022). End-of-study interviews identified that both patients and clinicians found the system quick and easy to use. For patients, in particular, the system provided reassurance about symptom experience and the feeling of being listened to. The clinicians largely viewed the system as feasible

XSL•FO
RenderX

and acceptable, and areas that were mentioned included the early management of symptoms and connectivity between patients and clinicians, leading to enhanced communication.

Conclusions: This study demonstrates that remote monitoring and management of symptoms of people with MPM using a mobile phone are feasible and acceptable. The evidence supports future trials using remote symptom monitoring to support patients with MPM at home.

KEYWORDS

Introduction

Malignant pleural mesothelioma (MPM) is an asbestos-related cancer that affects the pleura surrounding the lung. Approximately 2700 people are diagnosed with mesothelioma each year in the United Kingdom; of those, 89% have MPM, most of whom are men [1,2]. MPM may progress in different ways; some people may have a period of stability, while others have a more progressive illness with a limited prognosis [3]. Many patients experience several symptoms simultaneously in addition to a high burden of emotional and psychological distress from early in the illness [4-6]. The most commonly reported symptom of MPM is chest pain, followed by breathlessness, fatigue, weight loss, and cough; these are accompanied by a reduction in quality of life resulting in considerable need for support [7-11]. Early supportive care interventions are essential to those with MPM so that a good quality of life is maintained for as long as possible [12]. With a shift in care from hospitals to local community settings, many patients with MPM must engage in self-care activities at home to prevent or reduce the severity of symptoms and treatment side effects and must make important decisions such as when to contact health services. Use of technological devices such as smartphones and tablets enables the monitoring of symptoms and increases the capacity for self-care as well as enhancing communication with health professionals.

Remote symptom monitoring is now available to support patients, such as those with MPM, throughout their illness with the use of devices such as smartphones to remotely collect and send data to health care providers for diagnostic interpretation or monitoring purposes. Remote symptom monitoring has been proven to be effective in the management of people with cancer being monitored at home [13]. A study by Basch et al [13] found that people with advanced cancer who reported their symptoms to health professionals using tablets were less likely to visit accident and emergency or be hospitalized, remained on chemotherapy for longer, and had improved survival. Furthermore, the benefits of using remote symptom monitoring have been demonstrated in large reviews of people with cancer and a range of life-limiting illnesses [14,15].

A remote symptom monitoring system called the Advanced Symptom Management System (ASyMS) has been developed by the authors. A study of the use of remote symptom monitoring of patients with lung cancer found that ASyMS led to improved symptom management and enhanced communication with health professionals [16]. ASyMS is one of the most evolved and tested remote symptom monitoring

systems in the field [17-19]. In this article, we discuss a study where we adapted ASyMS for patients with MPM, called ASyMSmeso, designed to meet the needs for symptom monitoring to support patients with MPM at home.

Methods

Study Aim and Objectives

The overall aim of the study was to adapt ASyMS for people with MPM (ASyMSmeso) and determine the feasibility and acceptability of integrating ASyMSmeso into oncology care delivery.

Specific objectives of the study were to (1) explore the experiences and perceptions of people with MPM, their carers, and health professionals while using the ASyMSmeso system through semistructured interviews with people with MPM, their carers, and health professionals; (2) describe changes in patient-reported outcome measures (PROMs; symptoms, supportive care needs, technology acceptance) over time for people with MPM using the ASyMSmeso system (Sheffield Profile for Assessment and Referral for Care [SPARC], Technology Acceptance Model [TAM] measure for eHealth, Lung Cancer Symptom Scale - Mesothelioma [LCSS-Meso]); (3) explore the impact of ASyMSmeso on processes and organization of care and the workforce using semistructured interviews with health professionals; and (4) collect information such as response rates, recruitment rates, adherence to the intervention, compliance to the study using data automatically collected by the system on user interactions, and alerts to inform a future trial of ASyMSmeso.

Sample

The study aimed to recruit a purposive sample of up to 45 people with MPM from 4 clinical sites across England and Scotland. This sample size was considered acceptable for a feasibility study [20]. The inclusion criteria for the study were that patients had received a definitive diagnosis of MPM, were deemed by a clinician to be both physically and psychologically able to participate in the study, and were predicted to survive for at least 6 months. We recruited 18 patients to the study.

The ASyMSmeso Intervention

The ASyMSmeso intervention is outlined in Figure 1. The patients used their patient handset (Samsung Galaxy J3 mobile phone) to complete a daily symptom questionnaire (DSQ) accessed through an application pre-installed on the mobile phone by the research team. The patients were required to complete a DSQ using their handset every day and at any time

they felt unwell. Data from the DSQ were received by the ASyMSmeso secure server and processed by the predefined alerting algorithm. Any symptom or symptom combination meeting the algorithm's conditions for alerting automatically generated an alert on the dedicated mobile phone handset carried by the responding health professional at the patient's hospital. The handset played an audio attention prompt on receipt of the alert. The responding health professional then acted on the alert within the pre-agreed 24-hour timeframe. The health professional used a hospital desktop computer to view the patient's symptom alert report and DSQ responses through a secure web-based clinician's dashboard before contacting the patient directly via telephone or text message to advise them on how to manage their symptoms. For symptoms that could be self-managed by the patient, the health professional was able to offer self-care advice. The patient could access self-care information at any time via the electronic library (eLibrary) section of the ASyMSmeso application installed on their patient handset and were also able to view graphs of their symptom profile over time. The responding health professional closed the alert by updating the patient record using the clinician dashboard.

Figure 1. The Advanced Symptom Management System for patients with malignant pleural mesothelioma (ASyMSmeso) intervention process.
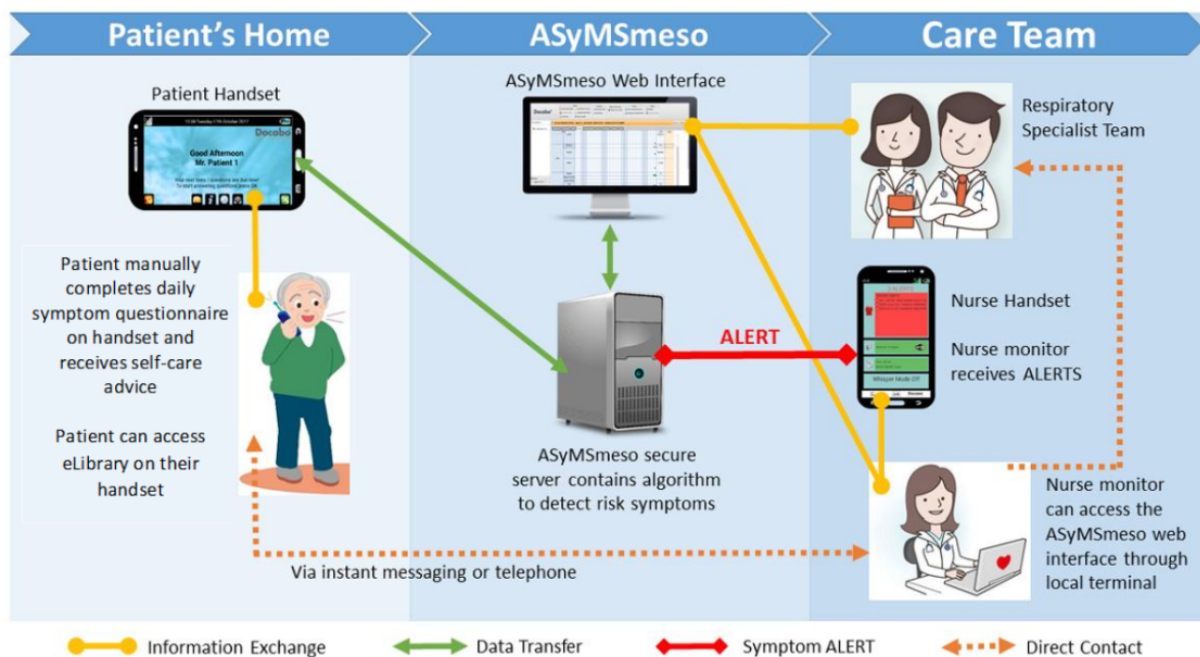
Procedure to Create the Daily Symptom Questionnaire and Alerting Algorithm

The symptoms included in the DSQ (breathlessness, pain, cough, sweating, fatigue, appetite, issues with indwelling pleural catheters, and constipation) and the alerting algorithm were informed by a literature search conducted by the research team and then refined with 3 focus groups with patients, carers, and health professionals (6 patients, 8 carers, 2 health professionals) before finally being reviewed by an expert panel of clinicians (3 respiratory consultants, 2 lung cancer nurse specialists, 2 mesothelioma nurse specialists) who agreed on the final contents of the DSQ and rules for the algorithm. It was agreed that daily monitoring of symptoms was appropriate as it enables early intervention at the start of the symptom trajectory, as demonstrated in our previous studies with ASyMS (16-18). The expert panel also agreed that an appropriate alert response time for a health professional responding to a patient with MPM is 24 hours.

Ethics

The ASyMSmeso intervention was delivered in addition to and complemented the patients' standard care. For any symptoms that required immediate medical attention, patients were asked to follow standard guidelines for their local area in the monitoring, management, and reporting of symptoms. Ethical approval was granted by West of Scotland Research Ethics Committee (reference number 17/WS/0077).

Data Collection

After providing informed consent, participants were asked to complete PROMs at baseline (T1), 6 weeks (T2), and the end of the study (T3). Midway through the study, we submitted an ethical amendment asking to reduce the period that people with MPM used the system, from 3 months to 2 months, to enhance recruitment within the remaining study timelines. For those patients participating in the study for 2 months, data were collected at baseline and the end of study (8 weeks) only. Patients completed the following PROMs: LCSS-Meso, SPARC, and TAM for eHealth.

LCSS-Meso

The LCSS-Meso is an 8-item patient scale that evaluates 5 domains, including overall symptomatic distress, functional activities, and global quality of life. Studies indicate it takes between 3 minutes and 8 minutes to complete [21].

SPARC

SPARC is a holistic needs assessment tool that covers physical, psychological, social, spiritual, and financial issues, with estimated completion times of <15 minutes [22,23].

TAM for eHealth

The TAM for eHealth questionnaire measures perceptual constructs from the information technology acceptance models: intrinsic motivation, perceived ease of use, perceived usefulness/extrinsic motivation, and behavioral intention to use eHealth. The original scale was revised from 20 to 12 items to keep it short and more practical for real-world situations. The TAM has been adapted for patient groups [24,25].

DSQs

In addition to the PROMs, we also analyzed the DSQs, and these formed part of the data analysis.

Demographic and Clinical Data

A health professional at the clinical site also completed a clinical and demographic questionnaire capturing patient data concerning age, gender, marital status, number and age of children, education level, occupation, diagnosis, stage of disease, length of time since diagnosis, treatment, Eastern Cooperative Oncology Group (ECOG) performance status, medications, and existing comorbidities (Table 1).

Table 1. Demographic characteristics of participants and their Eastern Cooperative Oncology Group (ECOG) status (N=18).

Characteristics	Results, n (%)
Age (years)	
55-64	4 (22)
65-74	8 (44)
75-84	5 (28)
Unknown	1 (6)
Gender	
Male	13 (72)
Female	5 (28)
Marital status	
Married	15 (83)
Widowed	1 (6)
Divorced	1 (6)
Co-habiting	1 (6)
Education	
Schooling incomplete	2 (11)
Finished high school	5 (28)
Further education	2 (11)
University	4 (22)
Trade qualification	1 (6)
Unknown	4 (22)
ECOG Status	
0 - Fully active	6 (33)
1 - Restricted but ambulatory	11 (61)
2 - Unable to carry out work activities	1 (6)

Semistructured Interviews and Focus Groups

At the end of the study, patients and/or carers and health professionals were invited to take part in semistructured interviews and focus groups to discuss their experience with the system.

Study Recruitment

A total of 54 people with MPM across the 4 clinical sites were approached to take part in this study. Of these, 18 patients consented and participated (response rate, 33.3%). Reasons for refusal included not wanting to deal with technology (2/36, 6%) and not wanting a daily reminder of symptoms (2/36, 6%); 32 (32/36, 89%) did not give a reason for not participating. Of the 18 participants, 14 were recruited to the 3-month protocol, and 4 were recruited to the 2-month protocol. Participants were predominantly male (13/18, 72%), with an average age of 71.6 years (see Table 1).

Data Analysis

Summary statistics are reported for each time point for each outcome in the 3 PROMs used in the study. Changes in PROMs from baseline (T1) to T2 and T3 were assessed using a Wilcoxon matched pairs test. Male participants were compared to female participants for changes in PROMs using a Mann-Whitney U test. The associations between the change in each PROM at T2 and T3 with age, education, and ECOG status was assessed using Spearman rank correlation tests. Statistical significance is reported at the 5% level.

All interviews were recorded digitally, transcribed, and analyzed using thematic analysis as advocated by Braun and Clark [26] using NVivo 11.4.1.1064. As with convergent mixed methods, researchers working on the project concurrently conducted the quantitative and qualitative elements, analyzed the two components independently, and interpreted the results together [27].

Results

Patient-Reported Outcome Measures (PROMs)

PROMs were available for 18, 15, and 9 patients at T1, T2, and T3, respectively.

The completion rate of the PROMs was over 90%: 97.2%, 100%, and 93.8% for SPARC at T1, T2, and T3, respectively; 100% for TAM and LCSS-Meso at all time points.

No significant changes in PROMs (Table 2) were observed, as measured on the LCSS-Meso, from T1 to T2 or T3 using the Wilcoxon matched pairs test (P=.293; Table 2). Even though this was a feasibility study and therefore not powered to detect statistical significance, we did note a statistically significant (P=.049) improvement from baseline to T2 (6-8 weeks) in psychological need on the SPARC scale, which suggests that completing a DSQ and using ASyMSmeso appear to provide psychological support to patients. Also, a statistically significant (P=.022) improvement was observed in the "Usefulness" domain of the TAM from T1 to T2, which suggests that patients found the ASyMSmeso system more useful over time. No statistically significant changes were identified at T3, but the low sample size at T3 renders analysis impractical.

Table 2. Patient-reported outcomes.

Measurement	T1[a], mean (SD, range)	T1, median (IQR)	T1-T2[b], median (IQR)	P value[c]	T1-T3[d], median (IQR)	P value[e]
SPARC[f]						
Physical	12.88 (8.87, 2.00-31.50)	12.53 (13.5)	2.00 (10)	.196	4.50 (7.25)	.123
Psychological/emotional	3.28 (2.89, 0.00-9.00)	3.00 (5.25)	1.00 (2.00)	.049	1.00 (2.5)	.572
Spiritual/religious	0.67 (0.97, 0.00-3.00)	0.00 (1.25)	0.00 (1.00)	.161	0.00 (1.00)	.102
Independence	1.11 (1.23, 0.00-4.00)	1.00 (2.00)	0.00 (0.00)	1.00	0.00 (1.50)	.180
Family/social	1.00 (0.89, 0.00-3.00)	1.00 (1.50)	0.00 (0.5)	1.00	0.00 (0.75)	1.00
Treatment concerns	1.24 (1.56, 0.00-6.00)	1.00 (2.00)	0.00 (3.00)	.755	0.00 (2.5)	.683
TAM[g]						
Ease of use	4.93 (1.08, 3.33-7.00)	5.00 (2.09)	-1.17 (1.67)	.196	–0.50 (2.00)	.352
Usefulness	5.94 (0.84, 4.00-7.00)	6.00 (1.1)	0.00 (3.4)	.022	–0.30 (3.4)	.671
LCSS-Meso[h]	23.40 (18.12, 2.25-63.75)	13.69 (31.38)	–0.38 (26.88)	.859	–0.25 (28.06)	.575

[a]T1: baseline.

[b]T2: 6 weeks or 8 weeks.

[c]Comparing the change in median values from T1 to T2.

[d]T3: end of the study.

[e]Comparing the change in median values from T1 to T3.

[f]SPARC: Sheffield Profile for Assessment and Referral for Care.

[g]TAM: Technology Acceptance Model.

[h]LCSS-Meso: Lung Cancer Symptom Scale-Mesothelioma.

No statistically significant associations were found in the relationships between demographic or clinical variables and PROMs other than between age and the change from T1 to T2 in family/social issues on the SPARC questionnaire (P=.036), with a positive correlation coefficient of 0.584, indicating that for lower age groups, social and family issues increased over time.

Adherence to the ASyMSmeso Intervention

Adherence to the ASyMSmeso intervention was explored. In total, 18 patients completed a total of 1343 DSQs over a total of 1334 days. There were more responses to DSQs than days because patients could complete them whenever they felt unwell or a symptom had changed. The compliance rate was 88.5%, with a median of 93.2% for completing DSQs. This rate of compliance is especially impressive given that the average number of days a person took part in the study was 82.8 days. This suggests that those taking part in the study found completing DSQs to be both feasible and acceptable.

Alerts

Of the symptoms assessed on the DSQ, breathlessness was the most common symptom to generate an alert, followed by pain. There was a large number of additional symptoms that triggered alerts for individual patients such as "runny nose," "stiffness in the hips," and "urine infection." The number and types of alerts that were generated by the system were analyzed; 7 patients did not generate any alerts during the study. The median number of alerts generated per patient was 3. Most patients generated an alert fewer than 10 times. One patient was responsible for a third of all alerts, with 52 alerts, and the 4 highest alerting patients were responsible for 116 of the 154 total alerts. This equates to one alert generated approximately every 8.33 days per patient. The highest alerting patient generated 1 alert per day during the study. The alerts occurred mostly in the early morning and early evening. As a result, there was a proportion of alerts that were out of hours (ie, after 5 pm), which warrants consideration for future implementation.

Health professionals had 24 hours to respond to the alerts generated by the system. Following an alert, patients were contacted mainly by telephone, and the most common response to the alerts was "advice as before," primarily for the 1 patient who alerted daily with the same set of additional symptoms. Other responses centered on advice for supported self-care and how to access community or hospital services.

User Experience

Participating patients, their carers, and clinicians were invited to take part in an end-of-study focus group or semistructured interview to explore their experiences and perceptions with using the ASyMSmeso system and identify how the implementation of ASyMSmeso impacted existing oncology services. The decision as to whether an interview or focus group was held depended on the availability of the participants.

Patient and Carer Experiences

A total of 8 patients and 3 carers were interviewed at the end of the study. All those recruited to the study were approached to take part in interviews. However, a number declined, and 1 person had died, giving a 47% response rate. Most participants were male (n=6), with an average age of 71 years. Two themes emerged from the interviews with patients and carers, namely (1) positive experiences and ease of use and (2) feelings of reassurance.

Regarding the theme of positive experiences and ease of use, in general, patients found the system very easy to use and quickly embedded it into their daily routine. They reported that it took them less than 5 minutes to complete the DSQ, and they usually completed it at the same time every day, often first thing in the morning. The system was described as "straightforward … and so simple to use" [Patient 01].

Very few issues with using the system were noted, and all participants felt that the DSQ was relevant to people with MPM:

> …my activity level had changed; my walking ability had changed because of the breathlessness. So yeah, these were relevant to what was happening to me at that point in time. [Patient 03]

Patient 03 felt that the symptom monitoring was relevant to what was happening to him in terms of his mobility and identified the link with his breathlessness.

Regarding the theme of feelings of reassurance, for many people, the sense of reassurance offered by the system was important, whether or not they were experiencing symptoms. One patient reported how "it was reassuring to know that you're feeling stable and not having symptoms" [Patient 09]. Patient 09 felt reassured that there was no change in his symptom profile, and he felt stable; therefore, no recent sign of deterioration was reassuring to him.

For some, knowing that someone was listening was important:

> I think what was good was the fact that you always felt that there was somebody else at the other end of the line who was listening to what you were saying. [Patient 01]

This was an interesting finding that this patient felt he was being listened to: a sense of connectivity that brought a level of reassurance to him.

This feeling of connectivity was also mentioned by Patient 02 who felt reassured that someone was "watching him."

Others spoke positively about responses to alerts, which resulted in their symptoms being effectively managed:

> I kept getting bothered with my hips, and she said "have you taken anything for it?" She'd actually phoned me up, and I had said no. She said, what about paracetamol, and I said I'll give it a go. So, I gave it a go, and it fixed it [laughs]. [Patient 03]

The patients reported little trouble in using the mobile phone to report symptoms; in fact, they found this experience reassuring, particularly when there was no change in symptoms over time and they could conclude that they were stable. The connectivity provided by the system was found to be reassuring to patients that someone was keeping an eye on them and responding to their data by calling them back and helping to sort out the problem.

Health Care Professionals' Experience

At the end of the study, 11 health professionals with many years' experience caring for patients with MPM (2 respiratory consultants, 4 clinical nurse specialists, 2 oncology nurses, 3 research nurses) took part in the interviews. From the interviews,

3 themes emerged: ease of use, symptom prevention and management, and enhanced communication.

Regarding the ease of use theme, similar to the feedback from the patients, the majority of health care professionals found the system acceptable and easy to use:

> *It was good. I found that really easy...* [HP 11]
>
> *Access to a smartphone device makes sense.* [HP 09]

They spoke about how, in general, they found the system easy to use and navigate. They also highlighted the benefit to both patients and themselves in achieving a better patient experience:

> *And, you know, badging this as a patient experience or patient experience improvement, and if there's improvement in that, then that's better for both the patient but also the treating clinicians.* [HP 08]

Some clinicians referred to the challenges of using remote monitoring concerning fears of technology and change:

> *You know, as always, it's fear of change, isn't it? Fear of technology, fear of change, fear of who is going to be responding to these alerts, and will it mean more work? So, I think that allaying those fears of change, getting buy-in from clinicians would probably overcome that.* [HP 06]

Some of the issues in the transition and ease of change to remote monitoring appeared unique to the health professionals, especially the fear of change, as expressed in the previous paragraphs. Interestingly, the patients did not identify any fears of using technology. The health professionals were fearful of the potential for increasing workload, handling and responding to alerts, and the need to work at allaying fears as part of the intervention.

Regarding the theme of improvement in symptom prevention, clinicians spoke about the positive clinical benefits of the alerting system — particularly its role in the early management of symptoms. One clinician described how signs of a chest infection were picked up, which then in turn led to "identification of disease recurrence" [HP 04].

Clinicians also spoke about how symptoms that may otherwise have been missed were picked up earlier:

> *Whereas maybe a patient wouldn't necessarily call us for that, but we can call them back because you are noticing some changes in symptoms...* [HP 04]

One clinician found the system useful as a means of prioritizing care to those patients who were reporting more symptoms than others about things that they were not previously aware of:

> *It made me think about the way I assess people and what I ask them over the phone. Because sometimes what they alerted, [it] wasn't something they told me about when I spoke to them or when I saw them in clinic. Like maybe they were reporting on that [DSQ] that they hadn't eaten for a couple of days; then, when you see them in clinic, it's like "oh yeah, I've got a diet. I'm eating really well. The wife's making me a dinner." So, it was... probing a wee bit better into what was actually going on.* [HP 02]

HP02 identified the benefits she experienced regarding her patient assessment and how the system was beneficial for probing patients further about their diet and nutrition, for example, and getting a better history of what was going on with the patient.

Another professional commented on how the data on the ASyMSmeso system regarding what symptom management interventions were used following an alert would give them a facility to look back and see what treatments were helpful for the patient in the past. This would prevent the future use of clinical management strategies that were already known to be of limited benefit to that person.

Regarding the theme of communication and connectivity, a few clinicians thought that those with MPM may not want to dwell on symptoms every day and would therefore not benefit from using the system — particularly if they had no symptoms at all:

> *I would think that, say "this just reminds me of the fact that I've got cancer. This doesn't help me."* [HP 05]

This finding is interesting in light of comments from patients, many of whom felt that even if they had no symptoms, filling in the DSQ and noting symptoms were stable resulted in feelings of positive wellbeing and reassurance.

Similar to the patients, clinicians also recognized the benefits of changing the mechanisms by which patients communicated with them — particularly the way that the mobile phone system did not rely on the patient contacting the professional directly but did this automatically — reporting issues in a timelier manner:

> *I think for the patients, it may mean that they don't have to make that phone call. Because they just do it on the device and it goes through...I think ultimately that is generally a better thing, if they can let us know what is going on sooner rather than later.* [HP 05]

This anticipatory and preventative approach to care was also mentioned by HP03 as a good approach to having "infrastructure to manage it and monitor. And identify patients who are alerting and have early intervention to prevent deterioration in the symptom." [HP03]

However, there was also some frustration that unchanging symptoms continued to alert. For example, if someone continued to report severe pain, the system would keep alerting until the symptom severity reduced to a lesser level. This led to some health professionals suggesting that the alerting algorithm needed adjusting and to be more refined at an individual level.

Regarding the flow of information available in the eLibrary, one health professional suggested that the eLibrary could be extended:

> *To cover things like travel insurance and all the other little things if they could use it almost as their own personal sort of reference thing.* [HP 08]
>
> *Some of the tips were good though, I think. That again, I'm not sure if the patients took on board that bit of it.* [HP 10]

Overall, the clinicians felt that the system was easy to use and enabled them to work together with the patient to improve their experience and manage symptoms in a timely manner. They also identified that the system enabled a preventative approach to symptom monitoring and could potentially enable early intervention. Unlike the patients, the clinicians identified some drawbacks to the system, such as leading to increased workload in managing alerts and feeling that patients might not want to be reminded of symptoms every day.

Suggested Improvements

While people with MPM and clinicians reported the ASyMSmeso system easy to use, there were suggestions to improve system function. These included having the option for an alarm to remind patients to complete the DSQ. The system as it is currently configured requires clinicians to log into a laptop or desktop computer to view patient symptom reports and to handle alerts (Figure 1). A number of health professionals suggested that it would make sense to put this functionality on the clinician's mobile device. Other suggestions included amending the alerting algorithm to prevent over-alerting for unchanging symptoms.

Discussion

Principal Findings

The results demonstrate that this model of ASyMSmeso, a technology-enabled remote symptom monitoring system using PROMs, is feasible and acceptable to this population and those caring for them. Importantly, the findings provide new insights into effective ways to enhance the delivery of supportive care to people with MPM now and in the future, to improve patient outcomes and wellbeing. Our study demonstrated that those with MPM found the ASyMSmeso system easy to use, found it more useful over time, kept using it consistently for prolonged periods (2-3 months), and could see the value of using this remote technology in the management of their symptoms. These findings are notable, as it is well recognized that adherence, ease of use, and perceived usefulness are key factors in determining the successful implementation and scaling up of digital health technologies in health care contexts [24,28].

Supporting such positive findings is also the strong sense of reassurance that people with MPM expressed as a result of using this system. The virtual line of connectivity between patients and clinicians created by the ASyMSmeso system resulted in patients feeling a sense of reassurance and safety — knowing that there was "someone" at the other end watching over their symptoms and offering assistance if required. Such experiences resonate with other studies exploring the use of technology for the provision of supportive care to people with cancer [16,29]. A review of 14 qualitative studies exploring the use of telecare by people with cancer reported similar feelings of safety and reassurance and perceived improved communication with their health professionals as a result of using such systems [15]. The findings therefore are in agreement with previous studies conducted in other cancer populations that demonstrate that the monitoring of patient-reported outcomes using electronic platforms and devices is feasible and acceptable and has a positive impact on care outcomes [13,17,30]. In our study, we

also found a complementary benefit identified by the clinicians in terms of ease of use and identification of a preventative approach to symptom management enabled by communication and connectivity in the system.

Furthermore, of note is the significant reduction in psychological needs observed from baseline to 6-8 weeks post-study. While the small number of participants recruited to the study and the focus on feasibility limit inferences made from this finding, it is possible that such reductions in psychological need may well be based on the positive perceptions and experiences of patients using the system. A trial evaluating electronic symptom reporting in people with metastatic cancer reported improvements in survival compared to standard care [13]. It may be well worth exploring whether such benefits related to survival transfer to people with MPM — particularly in light of the lack of curative treatments at the present time.

In terms of changes in practice, a number of clinicians spoke about how using the system made them change the way that they assessed symptoms and prioritized care. They recognized that, by remotely monitoring symptoms on a daily basis, they had a much more detailed picture of the individual's symptom experiences over time, which could positively inform decision making and selection of supportive care interventions. Clinicians also identified benefits in changing the way that patients communicated with them — moving from a system that relied on patients recognizing symptoms and making a conscious decision to contact them to a system that reviewed patient symptom reports and automatically triggered alerts. Clinicians spoke about the benefits of the DSQ reporting of symptoms, particularly for those patients who were reluctant to bother clinicians by calling them to ask for help. These findings are of value in light of studies reporting that people with cancer continue to delay reporting life-threatening symptoms to health professionals [31].

The number of alerts generated throughout the study was relatively small and manageable. However, a significant number of alerts was out of hours. This warrants further investigation in future studies to understand why patients reported outside of normal working hours and what this means in terms of service provision. Furthermore, for the purpose of this study, clinicians managed alerts within the confines of their current practice (eg, 9 am to 5 pm, 7 days a week).

The PROMs used in the study were quick and easy to complete, with the LCSS-Meso taking about 4 minutes to complete and the SPARC, a holistic needs assessment tool, also quick and easy to complete. Adherence with completing these tools, including TAM, was very good, indicating the suitability for use in future trials.

Limitations

Limitations of this study include that there was a decline in the PROM response rate across the study time points. This was due in part to research nurses not always being able to collect data at the agreed time points and in part as a result of the change from a 3-month to 2-month data collection protocol, which reduced data collection from 3 to 2 time points.

Participation rates were low for the study, and it is possible that some of those who declined to participate in the study were doing so because they did not want to think about their symptoms on a daily basis, a possibility also raised by some health professionals. However, only a small number of nonrespondents (n=2) explicitly gave this as a reason. It is also true that study participation rates can be very low among people with advanced cancer [32], and a number of personal, social, and systemic factors have been identified that have an impact on a patient's decision to participate [33].

Conclusion

In conclusion, the results of this study demonstrate, for the first time, that the remote monitoring and management of symptoms of people with MPM using ASyMSmeso are feasible and acceptable, and the evidence generated strongly supports future studies scaling up this digital intervention to a wider cohort of patients with MPM.

Conflicts of Interest

None declared.

References

1. Mesothelioma statistics. Cancer Research UK. 2020. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/mesothelioma#heading-Zero [accessed 2020-10-20]
2. Yates DH, Corrin B, Stidolph PN, Browne K. Malignant mesothelioma in south east England: clinicopathological experience of 272 cases. Thorax 1997 Jun;52(6):507-512. [doi: 10.1136/thx.52.6.507] [Medline: 9227715]
3. Parker C, Neville E. Lung cancer * 8: Management of malignant mesothelioma. Thorax 2003 Sep;58(9):809-813. [doi: 10.1136/thorax.58.9.809] [Medline: 12947146]
4. Arber A, Spencer L. 'It's all bad news': the first 3 months following a diagnosis of malignant pleural mesothelioma. Psychooncology 2013 Jul;22(7):1528-1533. [doi: 10.1002/pon.3162] [Medline: 22903737]
5. Hollen PJ, Gralla RJ, Liepa AM, Symanowski JT, Rusthoven JJ. Adapting the Lung Cancer Symptom Scale (LCSS) to mesothelioma: using the LCSS-Meso conceptual model for validation. Cancer 2004 Aug 01;101(3):587-595 [FREE Full text] [doi: 10.1002/cncr.20315] [Medline: 15274072]
6. Moore S, Darlison L, Tod AM. Living with mesothelioma. A literature review. Eur J Cancer Care (Engl) 2010 Jul;19(4):458-468. [doi: 10.1111/j.1365-2354.2009.01162.x] [Medline: 19832887]
7. Peake MD, Entwisle J, Gray S. Clinical Presentation, Radiological Evaluation and Diagnosis. In: O'Byrne K, Rusch V, editors. Malignant Pleural Mesothelioma. New York, NY: Oxford University Press; Jan 04, 2007.
8. Darlison L, Mckinley D, Moore S. Findings from the National Mesothelioma Experience Survey. Cancer Nursing Practice 2014 Apr 07;13(3):32-38. [doi: 10.7748/cnp2014.04.13.3.32.e1063]
9. Ball H, Moore S, Leary A. A systematic literature review comparing the psychological care needs of patients with mesothelioma and advanced lung cancer. Eur J Oncol Nurs 2016 Dec;25:62-67. [doi: 10.1016/j.ejon.2016.09.007] [Medline: 27865254]
10. Hughes N, Arber A. The lived experience of patients with pleural mesothelioma. Int J Palliat Nurs 2008 Feb;14(2):66-71. [doi: 10.12968/ijpn.2008.14.2.28597] [Medline: 18414321]
11. Chapman E, Hughes D, Landy A, Whale J, Saunders M. Challenging the representations of cancer pain: experiences of a multidisciplinary pain management group in a palliative care unit. Palliat Support Care 2005 Mar;3(1):43-49. [doi: 10.1017/s1478951505050078] [Medline: 16594194]
12. Guglielmucci F, Franzoi IG, Bonafede M, Borgogno FV, Grosso F, Granieri A. "The Less I Think About It, the Better I Feel": A Thematic Analysis of the Subjective Experience of Malignant Mesothelioma Patients and Their Caregivers. Front Psychol 2018;9:205 [FREE Full text] [doi: 10.3389/fpsyg.2018.00205] [Medline: 29515509]
13. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
14. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote Patient Monitoring via Non-Invasive Digital Technologies: A Systematic Review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
15. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer Survivors' Experience With Telehealth: A Systematic Review and Thematic Synthesis. J Med Internet Res 2017 Jan 09;19(1):e11 [FREE Full text] [doi: 10.2196/jmir.6575] [Medline: 28069561]

XSL•FO

RenderX

16.   Maguire R, Ream E, Richardson A, Connaghan J, Johnston B, Kotronoulas G, et al. Development of a novel remote patient monitoring system: the advanced symptom management system for radiotherapy to improve the symptom experience of patients with lung cancer receiving radiotherapy. Cancer Nurs 2015;38(2):E37-E47. [doi: 10.1097/NCC.0000000000000150] [Medline: 24836956]

17.   Maguire R, Fox PA, McCann L, Miaskowski C, Kotronoulas G, Miller M, et al. The eSMART study protocol: a randomised controlled trial to evaluate electronic symptom management using the advanced symptom management system (ASyMS) remote technology for patients with cancer. BMJ Open 2017 Jun 06;7(5):e015016. [doi: 10.1136/bmjopen-2016-015016] [Medline: 28592577]

18.   Johnston B, Maguire R, Kearney N. 4180 ORAL Using Mobile Phone Technology to Assess Symptoms in Patients Receiving Palliative Care – The Advanced Symptom Management System (ASyMS®-P). European Journal of Cancer 2011 Sep;47:S303 [FREE Full text] [doi: 10.1016/s0959-8049(11)71346-6]

19.   Aldiss S, Taylor R, Soanes L, Maguire R, Sage M, Kearney N, et al. Working in collaboration with young people and health professionals. A staged approach to the implementation of a randomised controlled trial. Journal of Research in Nursing 2010 Sep 17;16(6):561-576 [FREE Full text] [doi: 10.1177/1744987110380803]

20.   Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j.2002.384.doc.x] [Medline: 15189396]

21.   Hollen PJ, Gralla RJ, Liepa AM, Symanowski JT, Rusthoven JJ. Adapting the Lung Cancer Symptom Scale (LCSS) to mesothelioma: using the LCSS-Meso conceptual model for validation. Cancer 2004 Aug 01;101(3):587-595 [FREE Full text] [doi: 10.1002/cncr.20315] [Medline: 15274072]

22.   Ahmed N, Bestall JC, Payne SA, Noble B, Ahmedzai SH. The use of cognitive interviewing methodology in the design and testing of a screening tool for supportive and palliative care needs. Support Care Cancer 2009 Jun;17(6):665-673. [doi: 10.1007/s00520-008-0521-2] [Medline: 18982363]

23.   Hughes P, Ahmed N, Winslow M, Walters SJ, Collins K, Noble B. Consumer views on a new holistic screening tool for supportive and palliative-care needs: Sheffield Profile for Assessment and Referral for Care (SPARC): a survey of self-help support groups in health care. Health Expect 2015 Aug;18(4):562-577 [FREE Full text] [doi: 10.1111/hex.12058] [Medline: 23414548]

24.   Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]

25.   Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]

26.   Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

27.   Creswell JW, Pablo-Clark VL. Designing and Conducting Mixed Methods Research, 2nd Edition. Los Angeles, CA: Sage Publications, Inc; 2011.

28.   Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]

29.   McCann L, Maguire R, Miller M, Kearney N. Patients' perceptions and experiences of using a mobile phone-based advanced symptom management system (ASyMS) to monitor and manage chemotherapy related toxicity. Eur J Cancer Care (Engl) 2009 Mar;18(2):156-164. [doi: 10.1111/j.1365-2354.2008.00938.x] [Medline: 19267731]

30.   Mooney KH, Beck SL, Wong B, Dunson W, Wujcik D, Whisenant M, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546 [FREE Full text] [doi: 10.1002/cam4.1002] [Medline: 28135050]

31.   Oakley CM, Taylor C, Ream E, Metcalfe A. Patients making potentially life threatening delays in reporting neutropenic sepsis: Grounded theory study. JCO 2017 May 20;35(15_suppl):e21560-e21560. [doi: 10.1200/jco.2017.35.15_suppl.e21560]

32.   Lara PN, Higdon R, Lim N, Kwan K, Tanaka M, Lau DH, et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. J Clin Oncol 2001 Mar 15;19(6):1728-1733. [doi: 10.1200/JCO.2001.19.6.1728] [Medline: 11251003]

33.   Bell JAH, Balneaves LG. Cancer patient decision making related to clinical trial participation: an integrative review with implications for patients' relational autonomy. Support Care Cancer 2015 Apr;23(4):1169-1196. [doi: 10.1007/s00520-014-2581-9] [Medline: 25591627]

Abbreviations

ASyMS: Advanced Symptom Management System
DSQ: daily symptom questionnaire
ECOG: Eastern Cooperative Oncology Group
eLibrary: electronic library
LCSS-Meso: Lung Cancer Symptom Scale-Mesothelioma

MPM: malignant pleural mesothelioma
PROM: patient-reported outcome measure
SPARC: Sheffield Profile for Assessment and Referral for Care
TAM: Technology Acceptance Model

XSL•FO
RenderX

Original Paper

Email Consultations Between Patients and Doctors in Primary Care: Content Analysis

Helen Atherton[1], BSc, MSc, MPH, PhD; Anne-Marie Boylan[2], BSc, PhD; Abi Eccles[1], BSc, DPhil; Joanna Fleming[1], BSc, MSc, PhD; Clare R Goyder[2], BSc, BM BCh; Rebecca L Morris[3], BA, MA, MSc, PhD

[1]Warwick Medical School, University of Warwick, Coventry, United Kingdom

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[3]NIHR Greater Manchester Patient Safety Translational Research Centre, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Helen Atherton, BSc, MSc, MPH, PhD
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
United Kingdom
Phone: 1 +442476151405
Email: h.atherton@warwick.ac.uk

Abstract

Background: Increasingly, consultations in health care settings are conducted remotely using a range of communication technologies. Email allows for 2-way text-based communication, occurring asynchronously. Studies have explored the content and nature of email consultations to understand the use, structure, and function of email consultations. Most previous content analyses of email consultations in primary care settings have been conducted in North America, and these have shown that concerns and assumptions about how email consultations work have not been realized. There has not been a UK-based content analysis of email consultations.

Objective: This study aims to explore and delineate the content of consultations conducted via email in English general practice by conducting a content analysis of email consultations between general practitioners (GPs) and patients.

Methods: We conducted a content analysis of anonymized email consultations between GPs and patients in 2 general practices in the United Kingdom. We examined the descriptive elements of the correspondence to ascertain when the emails were sent, the number of emails in an email consultation, and the nature of the content. We used a normative approach to analyze the content of the email consultations to explore the use and function of email consultation.

Results: We obtained 100 email consultations from 85 patients, which totaled 262 individual emails. Most email users were older than 40 years, and over half of the users were male. The email consultations were mostly short and completed in a few days. Emails were mostly sent and received during the day. The emails were mostly clinical in content rather than administrative and covered a wide range of clinical presentations. There were 3 key themes to the use and function of the email consultations: the role of the GP and email consultation, the transactional nature of an email consultation, and the operationalization of an email consultation.

Conclusions: Most cases where emails are used to have a consultation with a patient in general practice have a shorter consultation, are clinical in nature, and are resolved quickly. GPs approach email consultations using key elements similar to that of the face-to-face consultation; however, using email consultations has the potential to alter the role of the GP, leading them to engage in more administrative tasks than usual. Email consultations were not a replacement for face-to-face consultations.

XSL•FO
RenderX

Introduction

Background

Increasingly, consultations in health care settings are conducted remotely using a range of communication technologies. Email allows for a 2-way text-based communication, occurring asynchronously. In general practice settings internationally, there has been varied adoption of email for consultation with patients. In Denmark, it is mandatory for all general practices to offer email consultations to their patients [1]. In the Netherlands, a survey demonstrated that 52.8% of practices offered email consultations [2]. Other countries have a less organized approach to the use of email for consultation. In the United Kingdom, just 6% of general practices reported offering email consultations to all patients; however, 21% of individual general practitioners (GPs) reported using email to have a consultation with a patient [3], reflecting the unstructured nature of this phenomenon in the UK general practice. Overall, email for consultation is not yet an embedded approach to having a consultation with a patient in general practice; a 2018 survey across 27 European countries reported that less than a fifth (19%) of GPs used email routinely to interact with patients about health-related issues [4].

Research has demonstrated that in the United Kingdom, where email consultation is not a mandatory offering and is not supported by 2-way secure encrypted email, email consultation is offered by GPs to selected patients and for selected problems, with GPs providing their own email addresses (National Health Service [NHS] or other addresses) to patients who are deemed to use this communication channel sensibly, not abusing the direct line to the doctor [5-7]. The selective approach to offering email interaction may reflect the concerns that GPs have about providing routine email consultations and the lack of protocol for processing email consultations [6].

These concerns include medico-legal implications and clinical safety issues about digital exclusion (eg, older patients, patients with low literacy) and increased workload via multiple contacts from and lengthy email exchanges with patients. UK general practice is funded by capitation; therefore, GPs are reimbursed not for individual consultations but according to patient numbers, regardless of the number of consultations conducted [1,6,8]. The potential for inequality is likely to be exacerbated when GPs are selective about who engages in email consultations. Despite this, the benefits of using email consultations have been described. Some patients appreciate the convenience and perceive improved access [1,5], and studies have shown that both patients and GPs feel that it promotes and sustains the doctor-patient relationship [1,6,8]. The written nature of the communication is seen as positive [6].

Quantitative evidence about the impact of email consultations on GPs' workload, safety, and legal risks is inconclusive [9,10]. Studies have explored the content and nature of email consultations to understand whether concerns and benefits may be realized. Most previous content analyses of email consultations in primary care settings have been conducted in the United States and have found no evidence to support clinicians' worries about receiving inappropriate and excessive numbers of emails [11-13], patients tended to use email for clinical rather than administrative enquiries [12,14], and emails often contained one or more problems [15]; however, it should be noted that in the United States, patients pay per consultation. Emails were most likely to be requests from patients, most commonly for medications or treatments [12,16]. An analysis of the content of email consultations in general practice in the Netherlands showed that this type of consultation was used for clinical issues, most popularly for psychological issues; when specific diagnoses were examined, the most common use of email consultation was for hypertension and diabetes [2].

There are numerous differences between the US and UK health systems, and to date, there has not been a content analysis of emails in UK general practice. Given the informal and unstructured approach to email consultations taken by GPs working in the United Kingdom [6], exploring the content of email consultations offers a way to examine how this type of consultation works where it is being used in practice and to see whether it leads to extended email exchanges and inappropriate content or safety issues in these instances.

Objectives

We aim to explore and delineate the content of selected consultations conducted via email in English general practice by conducting the first content analysis of email consultations between GPs and patients in a UK setting.

Methods

We sought to conduct a content analysis of email consultations between GPs and patients. Content analysis seeks to code, categorize, describe, and examine patterns in textual data. Content analysis takes an inductive approach to exploring data, not seeking to produce representative or generalizable findings but instead to identify concepts and patterns in the data and to produce numeric representations of various phenomena within the data.

The research team comprised 6 researchers. Each researcher has experience in mixed methods, qualitative research, and all work in primary care research; one of the researchers is a GP. The chief investigator has expertise in research on the use of email for consultation in general practice.

Setting

The study was conducted in 2 general practices in the South-Central region of England, United Kingdom.

Sampling of Participants

We purposively sampled 2 general practices for the use of email consultation. Participating practices had to have GPs who were using email to consult with patients. When this study was conducted, usage of email consultation was very low in the UK general practice, with just 6% of general practices offering it and 21% of GPs using it [3]. This limited the available types and numbers of practices that we could recruit from.

The 2 practices were selected such that they were different from each other with regard to the level of deprivation (measured using the Index of Multiple Deprivation 2015 [17]), practice

XSL•FO
RenderX

size (the number of patients registered with the practice and the number of doctors), and the number of GPs within the practice who used email to consult with patients.

We selected 2 GPs to participate in the study, one from each practice. We collected information on the age, gender, and year of qualification of the participating GPs to inform the findings.

Sampling of Emails

The emails adhered to a previously published and applied definition of an email consultation: a consultation intended to be a 2-way clinical communication between a GP and patient that was initiated by either party [6]. The content analysis was conducted on fully anonymized emails.

We used the number of email consultations collected in a previous primary care–based content analysis, which included 2 primary care professionals, as a guide (81 emails) [14] and aimed to collect 100 email consultations in total, with the option of obtaining more if needed as data analysis progressed. We continued the analysis until no new codes or themes occurred within the data set.

Data Collection

Data collection was retrospective. Data collection occurred from February 01, 2017. Once recruited into the study, the participating GPs were asked to identify their last 50 email consultations (comprising all emails in the consultation) and make those available to the research team.

We collected fully anonymized email consultations from each participating GP, with both direct and quasi identifiers removed. We asked the participating GPs to retain the time and date that the email was sent so that we could look at the time at which the emails were sent and/or received, the number of emails in the consultation, and over what time period this occurred. Although emails were anonymized, we asked the participating GPs to withhold any emails that contained sensitive information or content and to note down if this happened.

Participating GPs gave written informed consent to participate in the study. Informed consent from patients was not required to use anonymized data for research purposes; however, before the data were anonymized and passed onto us, each patient was emailed by the GP to inform them about the study and to give them the option to opt out of having their emails in the study. No patients opted out, and 2 sent messages of support.

Procedures were put in place to ensure the safe transmission of anonymized data. A member of the participating general practice staff anonymized the content, which included removing direct identifiers (personal names, email addresses, and telephone numbers) and quasi identifiers (ethnicity, gender, date of birth, postcode, and location information). The messages were transferred electronically using a secure NHS email address to a researcher who is a practicing GP and holds a secure NHS email account. The GP author rechecked that the data were fully anonymized before sending them to the chief investigator. Any attachments to the emails were removed, but GPs were asked to note down if there had been any attachments so that we could present these data.

We also collected contextual data to provide background to the sample; both participating GPs noted the age and gender of the patients from each email. These aggregate data were sent separately and were not linked to the email content.

We used a strict protocol in the event of poor care or illegal activity being exposed within any of the emails.

Ethical and Research Governance Approvals

Ethical approval was obtained from the University of Warwick Biomedical and Scientific Research Ethics Committee (reference number REGO-2016-1807), and research governance approval was granted by the UK NHS Health Research Authority.

Data Analysis

We used a parallel design approach to our content analysis, using both descriptive and normative approaches [18].

The descriptive approach allowed us to identify the visible and obvious components of the data and to examine the frequency of their occurrence. We used this approach to examine each email for the following elements:

- The number of patients and their aggregated characteristics
- The number of emails in each consultation exchange, the time it took to respond to each email, and the period between the initial and final emails
- The time of the day emails were sent and responded to, both by patients and GPs
- The nature of the email content (eg, administrative and/or clinical content)
- Use of attachments (eg, images, audio files, or other supplements)

To give an overview of the breadth of content across the data set, we categorized the email content by presenting condition, and this allowed for emails where there was more than one condition discussed.

These descriptive data provided context for the normative approach to analysis. A normative approach involves the development of themes, in this case, to explore the use and function of email consultation. In this part of the analysis, the frequency of occurrence of data is not important, and a single occurrence of data could comprise a code.

For the normative approach, we produced an initial coding frame devised by the chief investigator and another researcher. Initially, this involved both researchers reading each email, making notes, and meeting to discuss these. We then examined the coding frameworks used in previous content analyses in health care settings [11,12,19] and the existing qualitative literature on the use of email consultation [20], the delivery of the consultation in general practice settings [21], and patient safety to develop the coding frame further by looking for overlaps [22].

The chief investigator and another researcher then formally coded every piece of data independently using our finalized coding frame, with the researchers then meeting to discuss where codes had been changed, merged, or added as data were examined again. Our patient safety researcher and practicing GP/researcher then read all the data and checked a sample of

the coding samples to look for anything that could have been missed or that should have been added to the coding frame. Furthermore, 2 authors entered the coded data into NVivo (QSR International) qualitative analysis software, which was used to sort the data (see Multimedia Appendix 1 for a copy of the final coding frame).

In line with our normative approach, we developed high-level themes that categorized the content of the emails. We did this using conceptual mapping, adding each code to a visual map by hand and observing where they linked, until we identified recurring themes. The 2 researchers leading the analysis conducted this process together. However, throughout the interpretation stage, we held several meetings to ensure parity across the team; to enhance the validity of the analysis, we ensured that as a team, we agreed on the themes and their content. We were asked by the research ethics committee to use paraphrased examples from emails to illustrate the findings rather than direct quotations to ensure anonymity was maintained.

Results

We obtained 100 email consultations that occurred with 85 individual patients. From practice 1, we obtained 50 email consultations with 45 patients, and from practice 2, 50 email consultations with 40 patients. The email consultations comprised 262 emails in total. None of the participating GPs reported editing or removing any email consultation because they deemed that the content was too sensitive to share.

Participants

The participating practices were both in semirural (having both rural and urban characteristics) locations. Although they both had different deprivation scores, they were still in the least deprived deciles. Practice 1 had a smaller number of registered patients and 5 doctors, whereas practice 2 was much larger. In practice 1, just 1 GP offered email consultations to patients. In practice 2, there were 3 GPs offering email consultations. One GP from each practice participated in the study. Both participating GPs were male and had been using email consultation for 5 years with selected patients and offering patients only informal guidance on how to use it. Table 1 provides the details of the participating practices and the 2 participating GPs. Across both practices, 56% (48/85) patients engaged in the email consultations were male, and 39% (33/85) of the email consultations were with those in the age group of 40 to 59 years, with most patients being older than 40 years. When analyzed by practice, those who engaged in email consultations in practice 1 were older than those in practice 2. See Table 2 for a detailed breakdown of email user characteristics.

Table 1. Characteristics of participating practices and general practitioners.

Characteristics	Practice 1	Practice 2
Practice characteristics		
Location	Semirural	Semirural
Deprivation score (1=most deprived and 10=least deprived)	10	7
Number of patients registered at the general practice (as indication of practice size), n[a]	5000	14,000
Doctors in participating practice, n	5	10
GPs[b] using email consultation within the practice, n	1	3
GP characteristics		
Gender	Male	Male
Age (years)	35	39
Period for which the participating GP had been providing email consultation (years)	5	5
Offering guidance to patients on how to use email consultation	Informal only, from GP	Informal only, from GP

[a]Value rounded to ensure anonymity of practice maintained.

[b]GP: general practitioner.

Table 2. Characteristics of patients and/or caregivers engaging in email consultations.

Characteristics	Whole sample (n=85), n (%)	Practice 1 (n=45), n (%)	Practice 2 (n=40), n (%)
Sex			
Male	48 (56)	25 (56)	23 (57)
Female	37 (44)	20 (44)	17 (43)
Age (years)			
16-20	6 (7)	5 (11)	1 (2)
20-29	4 (5)	2 (4)	2 (4)
30-39	9 (11)	1(2)	8 (20)
40-49	19 (22)	8 (18)	11 (28)
50-59	14 (16)	9 (20)	5 (13)
60-69	12 (14)	5 (11)	7 (18)
70-79	16 (19)	11 (24)	5 (13)
80-89	3 (4)	2 (5)	1 (2)
≥90	2 (2)	2 (5)	0 (0)

Handling of Emails

Both GPs approached email consultation in different ways, choosing to process and store them in different ways, though with the same outcomes, that the emails were recorded in the patient record as a consultation that had occurred.

Email Consultations

There were 100 email consultations, with a total of 262 emails. Although we asked for email consultations that were 2-way, 15 email consultations that were sent to us comprised 1 email. Furthermore, 10 of those comprised just an email from the patient to the GP. We do not know if these were responded to via telephonic or face-to-face conversation; it is only known that there was no email response from the GP. Overall, 5 of the 1 way emails were from the GP to the patient. These 5 emails involved the transfer of information from the GP to patient, for example, a link to a website discussed during a face-to-face consultation, and there was no expected response.

Email consultations ranged from 1 email to a maximum of 9 emails in the conversation. The median number of emails in a consultation was 2. The number of days from the first email to the last email in the consultation ranged from 1 day to 1689 days. This was because 1 email consultation was reinitiated after a gap of 4 years, 7 months, and 15 days. With this anomaly excluded, the range is from 1 day to 351 days. The median number of days from the first to the last email was 3. GPs' response times ranged from <1 day to 35 days, and the median time taken by GPs to respond was 2 days.

We were able to obtain the timestamp for 164 of the 262 emails; for those without a timestamp, the corresponding information was not present in the data transferred to us by the GPs. Of these, most emails were sent between 6 AM and 6 PM. Overall, 3 emails were sent after midnight and before 6 AM, and these were all sent by the GP from practice 2. More emails were sent by the GPs (18 emails) in the evening than by patients (7 emails). See Table 3 for full details.

Table 3. Time of the day when the emails were sent.

Time period	Sent by general practitioners (n=118), n (%)	Sent by patients (n=144), n (%)
Overnight (12 AM-6 AM)	3 (2.5)	0 (0)
Morning (6 AM-12 PM)	28 (23.7)	43 (29.9)
Afternoon (12 PM-6 PM)	31 (26.2)	34 (23.6)
Evening (6 PM-12 AM)	18 (15.3)	7 (4.9)
Not recorded	38 (32.2)	60 (41.7)
Total	118	144

Content of the Email

Emails were used to discuss a wide range of complaints. Most email consultations contained clinical information (93/100); among these, several included some administrative content also (48/100). Very few email consultations were for administrative purposes only, with no clinical content (7/100). See Table 4 for

the range of presentations discussed during an email consultation. These were not mutually exclusive, and more than one presentation appeared in many email consultations.

A total of 13 email consultations included an attachment. Overall, 9 of the attachments were photos, 2 were documents containing monitoring data (blood sugar levels and blood

pressure), 1 was a fact sheet, and 1 was a form. In 11 email consultations, the contact was from carers on behalf of patients, 4 of which were about children and 7 about adults. In 13 emails, the GPs directly requested the patients to come in for a face-to-face consultation rather than continue the email consultation.

Table 4. Presentations discussed during email consultations.

Presentation	Examples of what was discussed
Cardiovascular	Blood pressure readings
Sexual health and family planning	Erectile dysfunction and contraception
Dermatology	Eczema
Gastroenterology	Constipation
Ear, nose, and throat	Labyrinthitis
Neurology	Brain injury
Orthopedic	Planned surgery
Palliative care	Relative in palliative care
Metabolic	Diabetes
Urological	Varicocele
Infectious disease	Lyme disease
Immunology	Allergic reaction
Preventive health	Vaccination
Medication	Side effects
Symptoms and advice	Fatigue
Administrative	Request for repeat prescription

Use and Function of Email Consultation

Within our data set, we identified 3 overarching themes: the role of the GP and email consultation, the transactional nature of email consultation, and the operationalization of email as a mode of consultation.

Theme 1: The Role of the GP and Email Consultation

Email consultations did not always follow the traditional model of communication used in general practice. As email offered a direct line to the GP, patients could use email consultations to circumvent the receptionist who would normally act as gatekeeper to appointment booking and conducting administrative processes:

> *Yes, they are due the vaccinations, book the appointment, and I'll order the vaccine.* [GP's response to patient's email]
>
> *Do nurses at the practice do cholesterol tests?* [Patient's enquiry to GP]

This took the GP into an administrative role, or they directed the patient back to the reception staff. Sometimes, email consultations provided a *halfway*, with the patients asking whether they needed a face-to-face consultation for their particular presentation or providing information ahead of a consultation. Normally, patients might not be able to reach the GP to ask these questions directly:

> *I am sorry to hear you are in pain. Let's discuss at the appointment you have booked in for Tuesday.* [GP's response to patient's email]

Patients were looking for clinical rather than administrative (receptionist) input on whether they needed a face-to-face consultation. This was time-saving for the patient but caused additional work for the GP, including undertaking the work of booking appointments:

> *I can see you tomorrow between 10 and 12 if convenient?* [GP's response to patient's email]

This also applied to cases where an email consultation led to a face-to-face consultation; the GP made an appointment, which took them outside of their role but saved time and effort for the patient. Sometimes it was the GP who made contact with the patient, offering reminders and scheduling care, indicating that for the GP, it was a choice to take on these tasks:

> *I think we planned to redo your bloods around now, shall I book you in?* [GP's email to patient]

Theme 2: Transactional Nature of Email Consultation

In our data set, patients were largely responsible for initiating an email consultation. This mirrors how patients normally seek consultations, with most emails being initiated by patients. Multiple email consultations were initiated from the same patients within the data set, and there was an example of a patient *picking up* and restarting an email conversation after a long period.

Emails functioned as a way for the patients to request an action, for example, a change in a prescription or to ask directly for a test result, circumventing the administrative processes already in place to allow patients to request these particular actions from a GP:

Please can you alter my dose to be lower for the next fortnight? [Patient's request to GP]

I was wondering about the results of the blood test I had in March? [Patient's request to GP]

Where GPs initiated emails, it was for obtaining and/or providing information following up from a face-to-face consultation, for example, sending a link to a website on contraception choices or updates:

If you let me know how the cream goes, I can always ask for a specialist opinion. [GP's email to patient]

Email consultation was not only a way to contact patients but was also used to devolve responsibility back to the patients. There was an example of a GP using email to send information to patients, sending it to the patient via email during a face-to-face consultation and ready for the patient to use right after the consultation. In one email, a GP referred to having tried to call the patient and using email as the secondary option, as it was not possible to reach the patient via telephone, thus leaving the responsibility for the contact with the patient.

Theme 3: Operationalization of Email as a Mode of Consultation

The style of writing differed between GPs and among patients. Across emails from both GPs and patients, the salutation "Dear" was most commonly used but valedictions included "Best wishes," "Many thanks," and "Kind regards." One GP always signed off an email with their first name, the other as "Dr [surname]." One GP was demonstrably less formal than the other, and patients were less formal in response.

The function of the email consultations was for health-related communication between a patient and their health care professional, in this case, a GP. However, some emails included non–health-related communication of a personal nature alongside this, for example, a patient thanking a GP for their help with a communication issue:

Thank you for taking the time to help me last Friday, I really appreciated you contacting the specialist. [Patient's email to GP]

The GPs applied some of the behaviors recognized as being key elements of general practice consultation and often employed several of these within one consultation. We observed signposting, negotiating with, and diagnosing the patient:

Please can you book in with the nurse for that. [GP signposting the patient]

I would be happy to refer you, but I don't think it is necessary at this point in time. [GP negotiating with the patient]

I think the likely cause of your symptoms is an infection. [GP diagnosing patient]

The most common behaviors were to engage with the patient and to apply safety netting:

Many thanks for this information, it is enough for me to make a decision. [GP engages with the patient]

Please come back and see me in a month, but if you have any issues before, then please get back in touch. [GP using safety netting]

Some consultations reverted to face-to-face consultations. An implicit working threshold was reached in the communication that led to the GP asking the patient to attend a face-to-face consultation. This was often where some form of physical and/or visual contact was required:

If you don't mind coming in so I can have a look. [GP reverting to face-to-face consultation]

Beyond clinical reasons for reverting to a face-to-face consultation, there were examples of boundary setting in relation to email consultation, whereby one lengthy email consultation culminated in the GP asking the patient to make an appointment rather than continue the conversation via email:

I think we need to book a telephone call please. [GP reverting to telephone consultation]

Potential explicit safety issues were linked to the medium; for example, a delayed response could be an issue for urgent enquiries. As email consultations are remote and asynchronous, there is potential for emails to be missed or not read, without this delay in communication being immediately apparent to the patient or GP. There were 4 examples of this within the data, which led to a delayed response from GPs:

This email went to my junk folder, apologies. [Example of potential safety issue]

I was on leave, so I have only just seen your email. [Example of potential safety issues]

We only observed potential safety issues and not any actual safety incidents.

Discussion

Principal Findings

Most patients who were engaged in email consultations with the participating GPs were aged over 40 years, and over half of the users were male. The email consultations were mostly short and completed in a few days. There were exceptions, with one long and detailed email conversation noted. Emails were mostly sent and received during the day, although GPs did send some responses in the evening, outside of their working hours. Emails were mostly clinical in content rather than administrative and covered a wide range of clinical presentations. Of the 100 email consultations, 13 led to face-to-face consultations and 13 included an attachment. The styles of writing differed between the 2 participating GPs, with one being more formal than the other. This reflects the wide variation among GPs in their style of working and may influence how the correspondence is received and understood by patients. The email consultations had many of the same characteristics as face-to-face consultations. GPs deployed a range of actions including signposting, negotiating, and engaging with the patient.

Limitations

There are limitations to our study, and the findings should be considered in light of these limitations. The participating GPs

XSL•FO
RenderX

selected the email consultations sent to us, and so we may have ended up with a different range of emails than if we set up a study which collected email consultations as they happened. However, several included consultations that were long or critical of the GP, and there were no email consultations excluded because of their sensitive content. We received the emails once they had been selected and anonymized; therefore, there may be details or nuances that were missing as a result, for example, missing timings. However, as the data were collected retrospectively, there was no opportunity for the Hawthorne effect [23] to influence the behavior of either GP or patient in conducting the email consultation.

Our sample was limited in size, and the participating GPs were male and recently qualified, working in areas of low deprivation. This was because of the low levels of email consultation provided when the study was conducted [3]. We had taken an exploratory approach to data analysis and did not intend to extrapolate our findings across the entirety of general practice. Future studies could sample more widely and select doctors and practices from a wide range of areas. We do not have detailed information about the patients. We do not know how many other types of consultations they had with the GP. We do not know how long they had been using email to consult with the GP nor how it was initiated. Therefore, this analysis is conducted without these contexts, which could have influenced the depth of interpretation of the data.

Outside of a formalized email consultation system, emails were harder to obtain and required extensive data protection approaches. Systems using portals allow for more readily extracted data [16]. They also include a broader patient population than where email is used informally with selected patients, and we acknowledge that our sample limits the findings.

Comparison With Prior Work

This is the first UK content analysis of email consultations, and our findings match closely with previous content analyses conducted in primary care settings in different countries. The patterns of usage that we observed in this study, with email consultations most likely to be requests from patients and concerning clinical rather than administrative contents, have also been demonstrated in previous content analyses in the United States [11-14,19]. We found that most email consultations were short, not sensitive in content, and did not include urgent content; these characteristics have also been demonstrated in previous content analyses carried out in countries other than the United Kingdom [11,14,19]. These commonalities are useful in determining the applicability of such studies about email consultation to the UK health system, indicating that there are lessons that the United Kingdom can learn from studies conducted in other countries. This is important given that email consultation use is likely to increase in the United Kingdom.

We identified a range of presentations by patients across the email consultations, with patients using email to consult about a wide range of issues, most commonly regarding medications and treatments. This matches the findings of other content analyses that looked at reasons for presentation using samples

of email consultations [2,12,15,16,19]. Beyond content analyses, 2 studies set in the UK general practice have quantitatively examined the reasons that patients gave when leaving a web-based request for an appointment with a GP, known as an *online consultation* and acting as the first email in a consultation [24,25]. The reasons given closely mirror the presentations identified in our study, with administrative requests and medication enquiries featuring alongside a wide range of conditions. This is interesting, as our study did not seek to quantify or provide representative data but had similar findings to studies that used quantitative methodologies. Our findings were also well aligned with those identified in a study that looked at reasons for engaging in face-to-face consultations [26]. This may indicate that the written nature of an email consultation does not necessarily affect what people consult about as compared with a face-to-face consultation; however, further research is needed to confirm this.

The themes identified in our analysis of the content of email consultations were also similar to themes identified in qualitative research studies focused on the use of email consultation. Our study was not an in-depth qualitative exploration and therefore did not yield the same amount of data at the same depth. However, here we note that our themes can be seen in other studies. A qualitative interview study with GPs and patients about email consultation usage in the UK general practice found that patients liked to use email consultation as an alternative form of access to the GP, bypassing the reception staff [6]. Our study examined the text within an email consultation, and within the text, we saw that the role of the GP changed as a result of patients using email consultation. In the same qualitative interview study, both patients and GPs performed boundary setting in their use of email consultation, and we observed this in text-based interactions, classifying this as the operationalization of email consultation [6]. In a different qualitative interview study of patient and GP perspectives of email consultation in Denmark [1], it was found that patients valued the information transfer capability of email, and we observed emails being used for this purpose. This triangulation of findings indicates that future studies could benefit from combining both text-based analysis and in-depth qualitative approaches. Beyond the methodological implications, it indicates that examining the text of email consultations may be a way for health care services to understand how email consultation is functioning.

Conclusions

In most instances where email is used to have a consultation with a patient in general practice, we observed that the consultation was short, clinical in nature, and resolved quickly. The GPs approached email consultation using an approach similar to that of face-to-face consultations; however, using email consultation had the potential to alter the role of the GP, leading them to engage in more administrative tasks than usual. Email consultation was not a replacement for face-to-face consultation for these GPs, but provided another way for the GPs and patients to engage with each other. On the basis of our analysis of their content, this form of consultation has the potential to provide an additional form of access for patients

and to allow GPs to deal with certain issues quickly and without   a face-to-face consultation.

Authors' Contributions

The authors have approved the authorship information. All authors contributed to the design and conduct of the study, and to the drafting of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding frame.
[DOCX File , 16 KB - jmir_v22i11e18218_app1.docx]

References

1. Hansen CS, Christensen KL, Ertmann R. Patients and general practitioners have different approaches to e-mail consultations. Dan Med J 2014 Jun;61(6):A4863. [Medline: 24947631]

2. Huygens MW, Swinkels IC, Verheij RA, Friele RD, van Schayck OC, de Witte LP. Understanding the use of email consultation in primary care using a retrospective observational study with data of Dutch electronic health records. Br Med J Open 2018 Jan 21;8(1):e019233. [doi: 10.1136/bmjopen-2017-019233] [Medline: 29358442]

3. Brant H, Atherton H, Ziebland S, McKinstry B, Campbell JL, Salisbury C. Using alternatives to face-to-face consultations: a survey of prevalence and attitudes in general practice. Br J Gen Pract 2016 Jul;66(648):e460-e466. [doi: 10.3399/bjgp16X685597] [Medline: 27215571]

4. Valverde-Albacete J, Folkvord F, Lupiáñez-Villanueva F, Hocking L, Devaux A, Altenhofer M. Benchmarking Deployment of Ehealth Among General Practitioners 2018. Publications Office of the EU. 2018. URL: https://op.europa.eu/en/publication-detail/-/publication/3ba2eb8b-5c07-11e9-9c52-01aa75ed71a1/language-en/format-PDF [accessed 2020-12-02]

5. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018 Apr;68(669):e293-e300. [doi: 10.3399/bjgp18X694853] [Medline: 29378697]

6. Atherton H, Pappas Y, Heneghan C, Murray E. Experiences of using email for general practice consultations: a qualitative study. Br J Gen Pract 2013 Nov;63(616):e760-e767. [doi: 10.3399/bjgp13X674440] [Medline: 24267859]

7. Atherton H. The potential of alternatives to face to faceface-to-face consultation in general practice, and the impact on different patient groups: a mixed methods case study. Health Serv Deliv Res 2018;6(20):-. [Medline: 29889485]

8. Dash J, Haller DM, Sommer J, Junod Perron N. Use of email, cell phone and text message between patients and primary-care physicians: cross-sectional study in a French-speaking part of Switzerland. BioMed Central Health Serv Res 2016 Oct 5;16(1):549 [FREE Full text] [doi: 10.1186/s12913-016-1776-9] [Medline: 27716256]

9. Atherton H, Sawmynaden P, Sheikh A, Majeed A, Car J. Email for clinical communication between patients/caregivers and healthcare professionals. Cochrane Database Syst Rev 2012 Nov 14;11:CD007978. [doi: 10.1002/14651858.CD007978.pub2] [Medline: 23152249]

10. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. J Am Med Assoc 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]

11. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]

12. Mirsky JB, Tieu L, Lyles C, Sarkar U. A mixed-methods study of patient-provider E-mail content in a safety-net setting. J Health Commun 2016;21(1):85-91 [FREE Full text] [doi: 10.1080/10810730.2015.1033118] [Medline: 26332306]

13. Abel EA, Shimada SL, Wang K, Ramsey C, Skanderson M, Erdos J, et al. Dual use of a patient portal and clinical video telehealth by veterans with mental health diagnoses: retrospective, cross-sectional analysis. J Med Internet Res 2018 Nov 7;20(11):e11350 [FREE Full text] [doi: 10.2196/11350] [Medline: 30404771]

14. Anand SG, Feldman MJ, Geller DS, Bisbee A, Bauchner H. A content analysis of e-mail communication between primary care providers and parents. Pediatrics 2005 May;115(5):1283-1288. [doi: 10.1542/peds.2004-1297] [Medline: 15867036]

15. Stiles RA, Deppen SA, Figaro MK, Gregg WM, Jirjis JN, Rothman RL, et al. Behind-the-scenes of patient-centered care: content analysis of electronic messaging among primary care clinic providers and staff. Med Care 2007 Dec;45(12):1205-1209. [doi: 10.1097/MLR.0b013e318148490c] [Medline: 18007171]

16. Shimada SL, Petrakis BA, Rothendler JA, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two veterans health administration medical centers: message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 1;24(5):942-999. [doi: 10.1093/jamia/ocx021] [Medline: 28371896]

17. English indices of deprivation 2015. Ministry of Housing, Communities and Local Government. 2015 Sep 30. URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [accessed 2020-12-02]

18. Bauer M. Qualitative Researching with Text, Image and Sound. London, England: AGE Publications Ltd; May 2000:132-152.

19. Sittig DF. Results of a content analysis of electronic messages (email) sent between patients and their physicians. BioMed Central Med Inform Decis Mak 2003 Oct 1;3:11 [FREE Full text] [doi: 10.1186/1472-6947-3-11] [Medline: 14519206]

20. Almond S, Mant D, Thompson M. Diagnostic safety-netting. Br J Gen Pract 2009 Nov;59(568):872-874. [doi: 10.3399/bjgp09X472971] [Medline: 19861036]

21. Howie JG, Heaney D, Maxwell M. Quality, core values and the general practice consultation: issues of definition, measurement and delivery. Fam Pract 2004 Aug;21(4):458-468. [doi: 10.1093/fampra/cmh419] [Medline: 15249538]

22. Makeham MA, Stromer S, Bridges-Webb C, Mira M, Saltman DC, Cooper C, et al. Patient safety events reported in general practice: a taxonomy. Qual Saf Health Care 2008 Feb;17(1):53-57. [doi: 10.1136/qshc.2007.022491] [Medline: 18245220]

23. Parsons HM. What happened at Hawthorne: new evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science 1974 Mar 8;183(4128):922-932. [doi: 10.1126/science.183.4128.922] [Medline: 17756742]

24. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. Br Med J Open 2017 Nov 22;7(11):e016901. [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]

25. Castenholz A. Scanning electron-microscopic study of Reissner's fiber in the caudal spinal cord of the cat and rabbit. Cell Tissue Res 1984;237(1):181-183. [doi: 10.1007/BF00229214] [Medline: 6478479]

26. Stuart B, Leydon G, Woods C, Gennery E, Elsey C, Summers R, et al. The elicitation and management of multiple health concerns in GP consultations. Patient Educ Couns 2019 Apr;102(4):687-693 [FREE Full text] [doi: 10.1016/j.pec.2018.11.009] [Medline: 30473249]

Abbreviations

GP: general practitioner
NHS: National Health Service

<u>Original Paper</u>

Impact of Patient-Clinical Team Secure Messaging on Communication Patterns and Patient Experience: Randomized Encouragement Design Trial

Stephanie L Shimada[1,2,3], PhD; Mark S Zocchi[1,4], MPH; Timothy P Hogan[1,5], PhD; Stefan G Kertesz[6,7], MD, MSc; Armando J Rotondi[8,9,10], PhD; Jorie M Butler[11,12], PhD; Sara J Knight[11,12], PhD; Kathryn DeLaughter[1], MA; Felicia Kleinberg[1], MSW; Jeff Nicklas[1,13], MSc; Kim M Nazi[14], PhD; Thomas K Houston[15], MD, MPH

[1]Center for Healthcare Organization and Implementation Research (CHOIR), VA Bedford Healthcare System, Department of Veterans Affairs, Bedford, MA, United States

[2]Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA, United States

[3]Division of Health Informatics and Implementation Science, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Heller School for Social Policy and Management, Brandeis University, Waltham, MA, United States

[5]Department of Population and Data Sciences, UT Southwestern Medical Center, Dallas, TX, United States

[6]Birmingham VA Medical Center, Department of Veterans Affairs, Birmingham, AL, United States

[7]Division of Preventive Medicine, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

[8]Center for Health Equity Research and Promotion (CHERP), VA Pittsburgh Healthcare System, Department of Veterans Affairs, Pittsburgh, PA, United States

[9]Center for Behavioral Health, Media and Technology, University of Pittsburgh, Pittsburgh, PA, United States

[10]Mental Illness Research Education and Clinical Center (MIRECC), VA Pittsburgh Healthcare System, Department of Veterans Affairs, Pittsburgh, PA, United States

[11]Innovation, Decision Enhancement & Analytic Sciences (IDEAS) Center, George E. Wahlen Veterans Affairs Medical Center, Salt Lake City, UT, United States

[12]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake, UT, United States

[13]Department of Social and Behavioral Sciences, University of California, San Francisco, CA, United States

[14]KMN Consulting Services, LTD, Coxsackie, NY, United States

[15]Section on General Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

Corresponding Author:
Stephanie L Shimada, PhD
Center for Healthcare Organization and Implementation Research (CHOIR)
VA Bedford Healthcare System
Department of Veterans Affairs
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 7816872208
Email: stephanie.shimada@va.gov

Abstract

Background: Although secure messaging (SM) between patients and clinical team members is a recommended component of continuous care, uptake by patients remains relatively low. We designed a multicomponent Supported Adoption Program (SAP) to increase SM adoption among patients using the Veterans Health Administration (VHA) for primary care.

Objective: Our goals were to (1) conduct a multisite, randomized, encouragement design trial to test the effectiveness of an SAP designed to increase patient engagement with SM through VHA's online patient portal (My HealtheVet [MHV]) and (2) evaluate the impact of the SAP and patient-level SM adoption on perceived provider autonomy support and communication. Patient-reported barriers to SM adoption were also assessed.

Methods: We randomized 1195 patients at 3 VHA facilities who had MHV portal accounts but had never used SM. Half were randomized to receive the SAP, and half served as controls receiving usual care. The SAP consisted of encouragement to adopt

XSL•FO
RenderX

SM via mailed educational materials, proactive SM sent to patients, and telephone-based motivational interviews. We examined differences in SM adoption rates between SAP recipients and controls at 9 months and 21 months. Follow-up telephone surveys were conducted to assess perceived provider autonomy support and self-report of telephone communication with clinical teams.

Results: Patients randomized to the SAP had significantly higher rates of SM adoption than the control group (101/595, 17.0% vs 40/600, 6.7%; $P<.001$). Most adopters in the SAP sent their first message without a motivational interview (71/101, 70.3%). The 10-percentage point difference in adoption persisted a full year after the encouragement ended (23.7%, 142/600 in the SAP group vs 13.5%, 80/595 in the control group, $P<.001$). We obtained follow-up survey data from 49.54% (592/1195) of the participants. SAP participants reported higher perceived provider autonomy support (5.7 vs 5.4, $P=.007$) and less telephone use to communicate with their provider (68.8% vs 76.0%, $P=.05$), compared to patients in the control group. Patient-reported barriers to SM adoption included self-efficacy (eg, not comfortable using a computer, 24%), no perceived need for SM (22%), and difficulties with portal password or login (17%).

Conclusions: The multicomponent SAP was successful in increasing use of SM 10 percentage points above standard care; new SM adopters reported improved perceptions of provider autonomy support and less use of the telephone to communicate with their providers. Still, despite the encouragement and technical assistance provided through the SAP, adoption rates were lower than anticipated, reaching only 24% at 21 months (10% above controls). Common barriers to adoption such as limited perceived need for SM may be more challenging to address and require different interventions than barriers related to patient self-efficacy or technical difficulties.

Trial Registration: ClinicalTrials.gov NCT02665468; https://clinicaltrials.gov/ct2/show/NCT02665468

Introduction

Secure messaging (SM) is a secure, asynchronous, patient-provider or patient-clinical team electronic communication channel that may help with care coordination and enable more efficient patient-provider interactions [1-3]. Patients use SM to ask questions or to keep their providers and clinical teams informed about their health status in between medical visits [4-6]. Several observational studies and multiple systematic reviews have found that SM can have a positive impact on health outcomes and patient satisfaction for some patients [6-12]. While some providers have expressed concern that SM may be difficult to keep up with and interrupt workflow, SM may help improve productivity by reducing telephone communication and improving visit efficiency [13].

The Centers for Medicare and Medicaid Service Meaningful Use requirement has made SM a common feature of patient portals and tethered personal health records across many health care systems [4,14-16]. In 2004, the Veterans Health Administration (VHA) implemented My Health*e*Vet (MHV), an online patient portal and personal health record that has an SM feature to support communication between patients and their VHA clinical team members [17,18].

However, the majority of VHA patients still do not use SM. As of September 2016, 12% of all VHA patients and 27% of MHV portal users were active users of SM who had sent at least one message to their clinical team in the previous 24 months. At the time of this writing (February 2020), active engagement with SM was at 16.7% of VHA patients. VHA patients tend to be older and face more complex health care needs than the general population and have less socioeconomic means than veterans that do not use the VHA [19]. Previously documented barriers

to SM adoption among veterans include lack of awareness about the SM features, not having a need for communication, limited access to technology, low computer literacy, and feelings that their provider does not support SM use [10,20,21]. Facilitators of SM adoption include understanding the purpose for using SM, having a health-related need that aligns with that purpose, and seeing SM as a convenient alternative form of communication [20].

SM has been found to be associated with positive health outcomes, such as improved HbA1c or blood pressure control, and improved antiretroviral adherence and HIV control in several observational studies [7,8,22-25]. However, these studies are prone to selection bias (ie, differential uptake of SM based on health status and other unobserved patient characteristics), and single-site, randomized trials are less generalizable. While research in this field is expanding rapidly, the evidence is not yet mature. Systematic reviews of SM have found many of the studies lacking in rigor [11,12]. To our knowledge, no multisite, randomized trial has tested an intervention to increase SM use or examined its impact on patient-reported outcomes. In practice, randomized trials of SM are difficult to conduct as patients randomized to use SM may never send an SM and control patients may decide to use SM on their own.

To address these challenges, we conducted a multisite, randomized, encouragement trial to test methods to improve patient adoption of SM and to better understand the benefits of SM adoption. Our hypotheses were that patients receiving encouragement to use SM would have higher rates of SM adoption, communicate through telephone less often, and perceive easier access to their provider compared to controls. Through follow-up telephone interviews, we examined barriers to SM adoption to better understand the reasons why patients

might be resistant to using SM, despite being encouraged to do so.

Methods

Design

As part of a VHA health care operations–funded quality improvement effort, we conducted a multisite, randomized, encouragement design trial, a design that is appropriate for situations when both control and intervention arms have the potential to access or use the treatment being evaluated [26]. Randomized, encouragement trials are a type of trial that retains many of the strengths of randomized clinical trials while allowing patients the opportunity and flexibility to engage in the intervention as they see fit [27]. In a randomized, encouragement trial, participants are randomized to receive encouragement to try the treatment or intervention of interest. It is expected that some participants randomized to receive the encouragement will not try the treatment and that some of the control participants will try the treatment on their own (2-sided noncompliance).

Figure 1. CONSORT diagram.
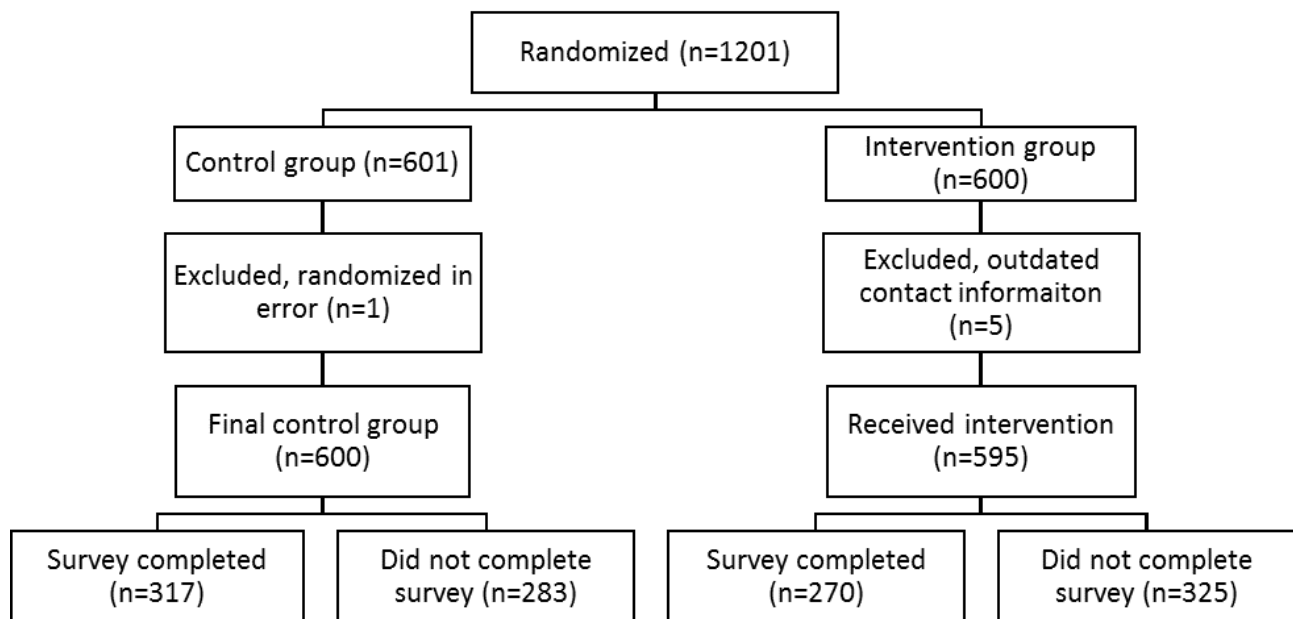
Setting

Three VHA medical centers participated in the trial. The sites were geographically diverse, located in the western, southern, and northeast regions of the United States, and included a relatively large (~30%) population of rural patients.

Participants

Veterans were eligible for the trial if they (1) had an authenticated account with the MHV online patient portal, (2) had never used SM, and (3) had a primary care appointment scheduled in the upcoming 2 months. Patients were excluded if they did not have a valid address or telephone number on record.

Randomization

Each of the 3 sites identified between 500 and 1200 patients who met the eligibility criteria, oversampling patients residing in rural zip codes. All sites used a block randomization table to assign 200 patients to the intervention (encouragement) arm and 200 as controls (Figure 1).

Intervention

Supported Adoption Program

Participants were randomized to be in the Supported Adoption Program (SAP) and receive encouragement to use SM to communicate with their clinical teams or to a control arm in which no encouragement was given. The SAP was comprised of multiple components, including 2 mailings, 2 SMs sent to participants from their primary care team's SM account, and 1 telephone-based motivational interview. The first mailing was sent in the first week of March 2016, and the last motivational interview was completed in September 2016.

The SAP components were developed to address key constructs of behavioral and motivational theories such as social cognitive theory (expectations, self-efficacy), theory of planned behavior (behavioral intention, subjective norms, attitudes), and the health belief model (perceived benefits, self-efficacy). The components were developed with input from MHV Coordinators, who regularly work with veterans to facilitate patient portal access and SM use. The mailings and team-initiated SMs highlighted reasons why patients like to use SM (such as convenience over telephone communication), how SM can benefit health, and assurance that their clinical teams want to communicate with them through SM. The first mailing contained a letter inviting participants to try SM, a brochure highlighting the advantages of SM, and a mousepad with step-by-step instructions on how to use SM. The first team-initiated SM was sent 3 weeks later and covered similar content as the first mailing. A second mailing was sent 3 weeks later and contained a letter reminding patients of the advantages of SM, a step-by-step instruction sheet for accepting the SM terms and conditions (for those unable to receive the provider-initiated SM), and a magnet with key telephone numbers for local and national support. A second team-initiated SM was sent to patients 2 weeks later with repeated content of the mailings.

Approximately 4 weeks after the second team-initiated SM, project staff at each site checked patients' portal activity and began making phone calls to patients in the SAP who had not yet sent an SM. The phone calls used motivational interviewing (MI) techniques to elicit behavior change by helping patients explore and resolve ambivalence or barriers to change [28-30]. The project staff were trained via telephone and given scripts for the motivational interviews. Three phone scripts were available based on whether the veteran (1) had read either of the SMs sent, (2) had not read the SMs sent, or (3) could not be sent the SM because they had not yet accepted the SM terms and conditions. Up to 5 attempts were made to reach each veteran. During the MI calls, project staff helped veterans troubleshoot common technical barriers during the call (eg, lost password) as well as other barriers to the use of SM (eg, perceived need, self-efficacy). Veterans were also provided instructions on how to contact their local MHV coordinator and the national MHV helpline.

Control Group

Participants randomized to the control group did not receive any component of the SAP.

Data Collection and Measures

Primary Outcome: SM Adoption

The primary outcome measured for this evaluation was SM adoption, which we defined as sending at least one SM during a 9-month period following the initial outreach (March 2016 to December 2016). We continued to monitor SM activity for an additional 13 months (until the end of 2017, when data collection ended) to measure longer-term effects of the intervention. SM activity data were collected via MHV data tables in the VHA corporate data warehouse.

Secondary Outcomes

Secondary outcomes were assessed via a telephone survey. The survey administration began approximately 6 months after initiation of the SAP. The survey asked veterans about the use of telephone communication with their health care teams, their perceptions of ease of communication with their health care teams, and perceptions of provider autonomy support, as measured by the Health Care Climate Questionnaire (HCCQ) [31], in the past 6 months. Those who had used SM during the evaluation period were also asked questions about their experience using SM. Patients that completed the survey received a US $20 gift card by mail.

The HCCQ is a series of 15 questions assessing perceived provider autonomy support (eg, "I feel my health care practitioner understands how I see things with respect to my health."). Patients answer on a scale of 1 (Strongly disagree) to 7 (Strongly agree). The HCCQ score is the average (mean) of these 15 items [31].

To measure the use of a telephone to communicate, patients were asked (yes or no), "In the past 6 months, have you communicated with your health care team by phone?"

To measure ease of communication, patients were asked, "How easy is it for you to communicate with your provider or health care team when you need to?" Patients that responded "easy" or "very easy" were considered to perceive easy access to their providers. Patients who responded, "very difficult," "difficult," or "neutral" were considered not to perceive easy access to their providers.

Self-reported barriers to SM adoption were collected during the motivational interviews. These barriers were coded into the following categories: low computer literacy (eg, not comfortable using a computer or navigating to the MHV website), difficulties with access (eg, no computer or internet), login difficulties (eg, lost password or username), effectiveness (eg, did not think they would get a reply), no perceived need for it, preference for in-person or phone communication, privacy concerns, and health-related barriers.

Secure Message Content

SMs sent by SAP and control patients during the 9-month evaluation period were double-coded for content by 2 research team members who met regularly to discuss and resolve differences in coding. Message content was coded into the following categories: Requests for Information, Requests for Action, and Information Sharing. The Requests for Information (eg, about symptoms, problems, medications, treatments) and Requests for Action (eg, requests for tests, medications, referrals) codes were based on a modified version of the Taxonomy of Requests by Patients [32,33]. The Information Sharing category captured messages that shared information about care obtained from other VA or non-VA providers, informational updates to the team about symptoms, vital readings (eg, blood pressure), or other personal or health-related topics. Previous studies have identified the need to share information with providers as one of the primary motivators of SM use [5,34].

Analyses

Descriptive statistics and chi-square tests were used to compare characteristics of patients randomized to the encouragement and control groups. We examined the success of the SAP by plotting the cumulative adoption rate over time for each group. Chi-square tests were used to compare adoption rates at 9 months and 21 months. Additionally, each component of the SAP (mailings, proactive SM, motivational interviews) was plotted graphically to examine adoption following each program component. Among patients that sent at least one SM, we used the Wilcoxson rank sum test to test for differences in the number of messages sent in each group. Chi-square tests were used to determine if the coded SM content differed between the groups. For patients that participated in a motivational interview, we identified common barriers to SM adoption and used a chi-square test to estimate the association of each barrier on later adoption. Finally, we compared characteristics of SM adopters to nonadopters regardless of treatment assignment using chi-square tests.

To estimate the effect of SM adoption on patient-reported outcomes, we conducted an intention-to-treat analysis (ITT), analyzing all participants with complete follow-up data as randomized. The ITT analysis compares outcomes in the SAP intervention group to the control group and does not consider whether or not patients sent an SM. ITT analysis is the preferred analytic approach for parallel-arm, randomized trials and more closely resembles real-life practice, where there is noncompliance of treatment. As such, the ITT analysis estimates the "effectiveness" of the SAP intervention rather than the "efficacy" of SM adoption. t tests were used for comparisons of continuous outcomes, and chi-square tests were used for comparisons of binary outcomes.

We also conducted a per-protocol analysis. A per-protocol analysis compares patients in control and intervention groups that completed the treatment as originally allocated. (ie, "compliers") [35]. For the SAP intervention group, this included all patients that sent an SM during the 9-month evaluation period (n=101). For the control group, this included all patients that did *not* send a secure message (n=560). Due to likely differences between compliers in each group, we used multivariable regression analysis, controlling for age, race, gender, marital status, rural residence, and copayment exemption. We used linear regression for continuous outcomes and logit regression for binary outcomes. All analyses were conducted using Stata v.15.1 (Stata Corp, College Station, TX).

Results

Across the 3 sites, 595 veterans were randomized to receive encouragement, and 600 served as controls. Veteran demographic and socioeconomic characteristics were balanced across the encouragement and control arms (Table 1). Approximately 30% (356/1195, 29.79%) of the veterans resided in rural areas, and 21.59% (258/1195) qualified for copayment exemptions based on their economic means. Approximately 21% (256/1195, 21.42%) of veterans were under the age of 50 years, 9.96% (119/1195) were female, 55.90% (668/1195) were married, and 19.83% (237/1995) were African American.

Table 1. Baseline characteristics of patients registered in the patient portal in 2016 who had not yet used secure messaging, with comparisons between the Supported Adoption Program (SAP) and control groups.

Characteristic	Control group (n=600), n (%)	SAP group (n=595), n (%)	Difference, %	χ^2 statistic	*P* value
African American	112 (18.7)	125 (21.0)	2.3	1.030	.31
Female	61 (10.2)	58 (9.7)	–0.5	0.058	.81
<50 years old	132 (22.0)	124 (20.8)	–1.2	0.239	.62
Married	341 (56.8)	327 (55.0)	–1.8	0.426	.51
Rural residence	172 (28.7)	184 (30.9)	2.2	0.728	.40
Copay exempt	132 (22.0)	126 (21.2)	–0.8	0.120	.73

Adoption of Secure Messaging at 9 and 21 Months, Compared Between the SAP and Control Groups

Veterans that received encouragement were more likely to send an SM than the controls. By the end of 9 months, 17.0% (101/595) of the veterans in the SAP had sent an SM compared to 6.7% (40/600) of the controls (*P*<.001). At 21 months, 23.7% (142/600) of veterans in the SAP had sent an SM compared to 13.5% (80/595) of the controls (*P*<.001; Figure 2).

Each component of the SAP (mailings, proactive SM, motivational interviews) was examined separately to explore the contribution of each on the overall rate of adoption (Figure 3). Of the 595 veterans in the encouragement arm who were sent the initial mailing, 17 (2.9%) sent an SM without any additional encouragement. In addition, veterans in this arm were sent 2 proactive encouragement SMs from their primary care team's SM account and an additional mailing. One-third (198/595) of the veterans could not receive the SMs because they never opted-in to receive secure messages. Of those that received the proactive SMs, 28.0% (111/397) opened and read the messages. This additional encouragement yielded 36 new users for a total of 53/595 (8.9%) new SM adopters before motivational interviews. Of the 542 veterans that were eligible for a motivational interview, 383 interviews were completed, and 30 of those that completed the interview sent an SM (30/383, 7.8%). An additional 18 veterans that could not be reached for the motivational interview (18/159, 11.3%) eventually sent an SM for a total of 101 (101/595, 17.0%) SM adopters at the end of the 9-month intervention.

XSL•FO

RenderX

Figure 2. Cumulative rate and point prevalence of secure message adoption among those randomized to the Supported Adoption Program (SAP) and controls.

Point Prevalence Adoption - Percent of Patients Who Sent a Secure Message		
	9 months	21 months
Control	6.7%	13.5%
Supported Adoption Program	17.0%	23.7%
SAP vs controls	10.3%	10.2%
	P<.001	*P*<.001
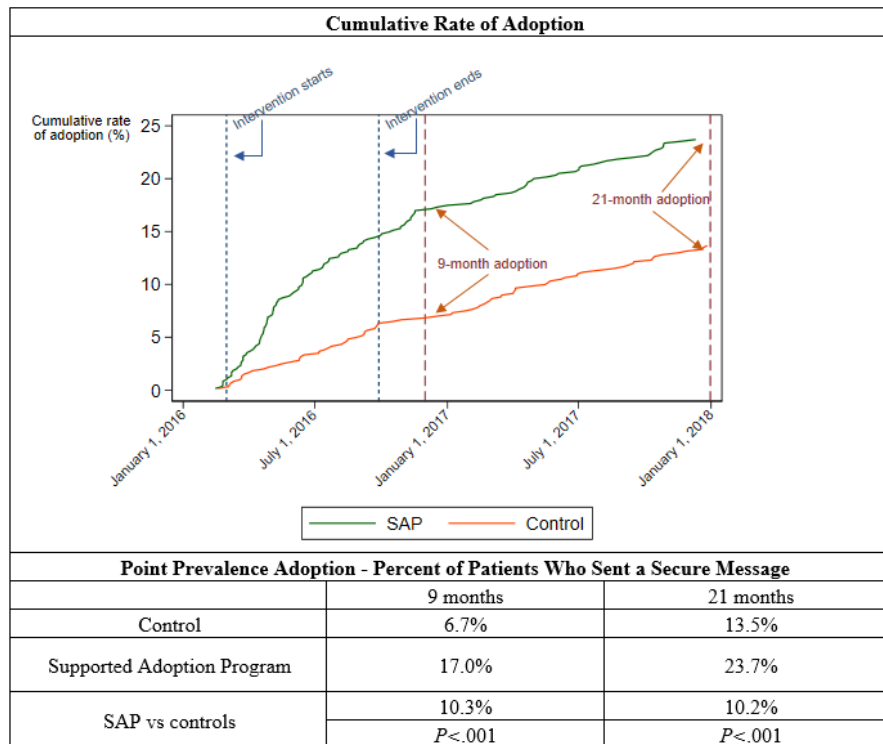
Figure 3. Adoption after each component and cumulative adoption over 9 months of active intervention among patients randomized to the Supported Adoption Program. SM: secure message.
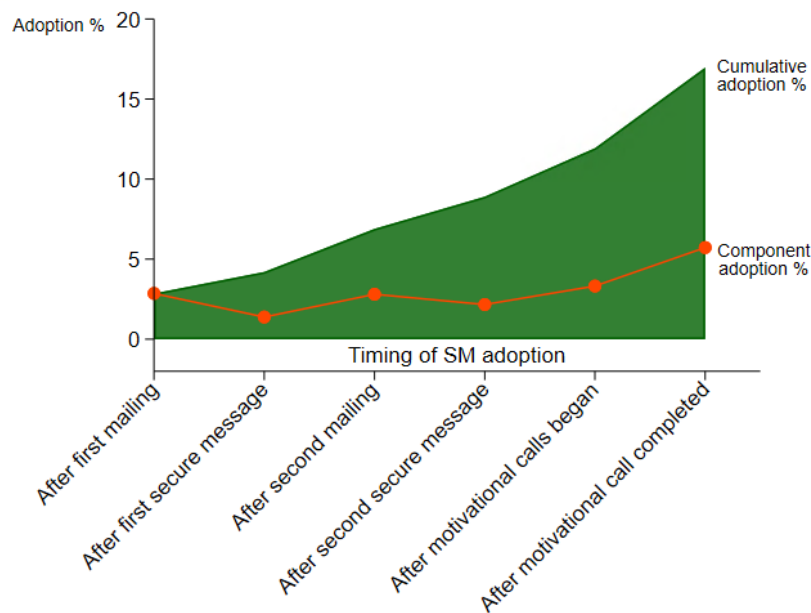
SM adopters were more likely than nonadopters to be female and younger than 50 years (see Multimedia Appendix 1) We examined the total number of SMs and content of the SMs sent by veterans. A total of 443 new messages were sent during the 9-month evaluation period. Among those who sent an SM, there was no difference between the encouragement and control groups in the average number of messages sent (*P*=.31). SM adopters in both groups sent an average of 2 messages (IQR 1-3). The content of the SMs that were sent by patients was also examined to determine if those in the encouragement arm communicated with their providers differently than the controls. The majority of SMs sent by patients in both groups were information requests or action requests about medications or treatments. We did not find any difference between the 2 groups in the frequency of these types of messages; however, controls were more likely to write to share information about their vitals or provide updates on symptoms or care obtained from other VA or non-VA providers (Multimedia Appendix 2).

During motivational interviews, patients were asked whether they were experiencing any significant barriers to SM adoption. Commonly reported barriers included low self-efficacy (eg, not comfortable using a computer, 24%), no perceived need for SM (22%), and difficulties with portal password or login (17%). The barrier most associated with non-adoption was having portal password or login difficulties (X^2_1=9.395, P=.002). Only 1 patient that identified this as a barrier eventually sent an SM (1/66, 1.5%). Patients that reported at least one significant barrier were much less likely to adopt SM than those that did not identify any (6.5% vs 25.0%, P<.001; Multimedia Appendix 3).

Impact of Secure Messaging on Perceived Health Care Climate and Communication

Table 2 shows the results of the ITT and per-protocol analyses for the 3 follow-up outcomes. Compared with the control group,

patients that received encouragement were less likely to report using the phone to communicate with their providers (68.8% [SAP group] vs 76.0% [control group], P=.05) and were more likely to perceive their provider as autonomy supportive (5.7 [SAP group] vs 5.4 [control group] on a 7-point scale, P=.006). Ease of communication did not significantly differ between the two groups. In the per-protocol analysis, SAP recipients that adopted SM were less likely to report using the phone to communicate with their providers compared to the control patients that did not adopt (59.2% vs 77.1%, P=.003). Provider autonomy support and ease of communication did not significantly differ between SAP adopters and control nonadopters.

Table 2. Impact of the Supported Adoption Program (SAP) and secure message adoption: intention-to-treat (ITT) and per-protocol analyses.

Outcome measure	ITT				Per-protocol			
	SAP (n=270)	Control (n=317)	Difference (95% CI)	P value	SAP adopters (n=51)	Control non-adopters (n=293)	Difference (95% CI)	P value
Health care climate scale, mean (SD)	5.7 (1.1)	5.4 (1.3)	0.3 (0.1 to 0.5)	.006	5.7 (0.9)	5.4 (1.3)	0.3 (–0.1 to 0.7)	.18
Communicated by phone, n (%)	186 (68.8)	241 (76.0)	–7.2 (–14.5 to 0.0)	.05	30 (59.2)	226 (77.1)	–20.8 (–35.4 to –6.3)	.003
Easy to communicate, n (%)	176 (65.1)	187 (58.9)	6.2 (–1.7 to 14.1)	.12	36 (70.0)	173 (59.0)	11.3 (–2.5 to 25.1)	.13

Discussion

Principal Findings

In this randomized trial encouraging SM use, the SAP (encouragement intervention) resulted in an increase in SM use compared with control patients that received usual care. The SAP had a modest but significant impact on overall adoption, with 17% using SM, compared with 7% in the control arm at 9 months, and 24% versus 14% at 21 months.

The majority of new adoption resulted following the low-intensity mailings and SM components. While the motivational interviews resulted in 30 new SM adopters, most new users did not participate in a motivational interview. The mailings or SM components of the SAP would be relatively inexpensive to implement more broadly. The 2 mailings cost approximately US $1.50 to US $5.00 each (depending on whether a small magnet or mousepad was included) and would be scalable to larger groups of patients. We estimate that it would take 5 minutes to send a templated SM with encouragement content to new patient portal users. The motivational interviews took about 30 minutes each to complete and were estimated to cost US $15.50 per completed call based on the average salary of those trained to make the calls. This suggests that, for relatively low effort, health care facilities could engage patients via mailings or SMs highlighting the benefits of SM and addressing barriers to SM use. This type of

outreach may also be effective to encourage adoption of other new patient-facing technologies.

In addition to evaluating the program's effectiveness at getting veterans to use SM, we monitored its impact on self-reported outcomes. Our ITT analysis revealed an increase in perceived provider autonomy support for the encouragement arm compared with the control arm. Increased perceived autonomy support has previously been shown to be associated with improved patient self-efficacy and health management behaviors, and proactive SM from health care teams to patients has been associated with greater perceived autonomy support [36]. It is possible that even templated encouragement to use SM sent from the primary care provider's team account may be adequate to achieve this benefit. Further, receiving the encouragement resulted in a shifting of patient-clinical team communication modality with a 7% lower rate of self-reported telephone contact in the encouragement arm compared with controls and a 20% lower telephone contact rate among SM adopters in the encouragement arm compared with control nonadopters. Although these are self-reported data, they suggest that SM does not necessarily increase overall communication with providers and may supplement other forms of communication. Future work should evaluate impact on telephone use more objectively.

We analyzed the content of the SM to see whether there were differences in the types of messages sent by those in the

encouragement group. Consistent with prior observational studies of SM content, information requests or action requests about medications or treatments were most common [5]. Our content coding did not reveal any significant differences in terms of frequency of requests for information or action; however, control patients were more likely to share information. This may in part be due to relatively small numbers of messages sent and the short timeframe for analysis. Patients may need to become more comfortable with messaging their provider before feeling ready to engage in more complex exchanges. However, it does suggest to us that the encouragement did not cause patients to send meaningless messages simply for the sake of sending an SM and that most patients will wait until they have a clinically relevant reason to SM their clinical teams.

Despite the demonstrated impact of the SAP, the majority of patients still did not engage in SM with their clinical team. We estimated that only one-quarter of those who were sent an SM opened and read the proactive SM. As suggested in the literature, additional encouragement from clinical teams might be needed to increase patient review of SMs [21]. In the VA, veterans are required to set preferences that determine whether they receive email alerts when there is a new SM in their patient portal. Patients who set their patient portal preferences to alert them to new incoming SM are more likely to read them [10]. Veterans who do not get an email alert may not even realize that they have been sent an SM. At the time of this trial, some veterans were unable to receive SMs because they had not accepted the feature's terms and conditions. This additional step is no longer required as the terms and conditions have been incorporated into those of the portal. Encouragement sent via SM may not be completely effective in systems where patients do not automatically receive message alerts to their regular email accounts or where there are other barriers to receiving messages.

Limitations

Both encouragement and control arms consisted of existing patient portal registrants. This type of encouragement program may be less effective if targeted to unregistered patients. Further, new users were disproportionately female and under 50 years old, compared with those who did not adopt SM. Thus, additional work may be needed to engage specific subpopulations.

The motivational interviews were conducted over a longer time period due to challenges reaching patients over the phone. As such, some patients that completed calls towards the end of the 9-month evaluation period had less follow-up time than those who completed calls earlier. When we followed all patients for an additional 13 months, SM adoption rates increased overall, but adoption remained consistently higher in the intervention group, as evidenced in Figure 2. Motivational interviews were conducted at each site by different project staff who were jointly trained in MI and used the same telephone scripts. We did not find differences in adoption rates by site, but it is still possible that different project staff may have been more or less effective than others at encouraging participants to start using SM.

While we monitored SM use beyond the 9-month follow-up period, we did not code additional messages sent by SM users after their initial months of use. Additional content coding of messages could have helped to determine whether the SAP may have shaped the content of their messages in any way over time.

Conclusions

This randomized, encouragement trial demonstrated that an SAP consisting of low levels of outreach to patients to help address known barriers to adoption of SM can successfully increase use. Some patients required more intensive support to begin use; however, our results show that over half of those who began SM use did so without a motivational phone call. About 70% (71/101) of patients that adopted SM did so without a motivational interview, suggesting that more limited outreach without motivational calls would still be effective at increasing SM adoption, while costing substantially less. Receiving information on SM and encouragement to use it had a positive impact on perceived provider autonomy support and, among SM adopters, resulted in lower self-reported use of telephone communication. Importantly, there were no negative impacts on frequency of SM communication or SM content, when compared with control patients who began using SM of their own accord. Low-intensity outreach can successfully engage patients in use of SM, and such use is associated with beneficial secondary outcomes such as improved perceived autonomy support and lower self-reported telephone communication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of secure messaging adopters.
[DOCX File , 14 KB - jmir_v22i11e22307_app1.docx]

Multimedia Appendix 2
Secure message content.
[DOCX File , 13 KB - jmir_v22i11e22307_app2.docx]

Multimedia Appendix 3
Barriers to adoption.
[DOCX File , 14 KB - jmir_v22i11e22307_app3.docx]

References

1. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273 [FREE Full text] [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]

2. Otte-Trojel T, de Bont A, Rundall T, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]

3. Hefner JL, MacEwan SR, Biltz A, Sieck CJ. Patient portal messaging for care coordination: a qualitative study of perspectives of experienced users with chronic conditions. BMC Fam Pract 2019 May 03;20(1):57. [doi: 10.1186/s12875-019-0948-1] [Medline: 31053063]

4. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]

5. Shimada S, Petrakis B, Rothendler J, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 01;24(5):942-949. [doi: 10.1093/jamia/ocx021] [Medline: 28371896]

6. Haun JN, Patel NR, Lind JD, Antinori N. Large-Scale Survey Findings Inform Patients' Experiences in Using Secure Messaging to Engage in Patient-Provider Communication and Self-Care Management: A Quantitative Assessment. J Med Internet Res 2015 Dec 21;17(12):e282 [FREE Full text] [doi: 10.2196/jmir.5152] [Medline: 26690761]

7. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]

8. Chung S, Panattoni L, Chi J, Palaniappan L. Can Secure Patient-Provider Messaging Improve Diabetes Care? Diabetes Care 2017 Oct 14;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]

9. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient Use of Electronic Prescription Refill and Secure Messaging and Its Association With Undetectable HIV Viral Load: A Retrospective Cohort Study. J Med Internet Res 2017 Feb 15;19(2):e34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]

10. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 06;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]

11. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]

12. Goldzweig CL, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, et al. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record. Washington (DC): Department of Veterans Affairs (US); Jul 2012.

13. Liederman EM, Lee JC, Baquero VH, Seites PG. Patient-physician web messaging. The impact on message volume and satisfaction. J Gen Intern Med 2005 Jan;20(1):52-57. [doi: 10.1111/j.1525-1497.2005.40009.x] [Medline: 15693928]

14. Kannry J, Beuria P, Wang E, Nissim J. Personal health records: meaningful use, but for whom? Mt Sinai J Med 2012 Sep 13;79(5):593-602. [doi: 10.1002/msj.21334] [Medline: 22976365]

15. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient Race/Ethnicity and Shared Medical Record Use Among Diabetes Patients. Medical Care 2012;50(5):434-440. [doi: 10.1097/mlr.0b013e318249d81b]

16. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao SR, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014 Dec 12;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]

17. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov 12;26 Suppl 2(S2):628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]

18. Chumbler NR, Haggstrom D, Saleem JJ. Implementation of Health Information Technology in Veterans Health Administration to Support Transformational Change. Medical Care 2011;49:S36-S42. [doi: 10.1097/mlr.0b013e3181d558f9]

19. Eibner C, Krull H, Brown K, Cefalu M, Mulcahy AW, Pollard MS, et al. Current and Projected Characteristics and Unique Health Care Needs of the Patient Population Served by the Department of Veterans Affairs. RAND Corporation. 2015. URL: https://www.rand.org/pubs/research_reports/RR1165z1.html [accessed 2020-10-15]

20. Haun JN, Lind JD, Shimada SL, Simon SR. Evaluating Secure Messaging From the Veteran Perspective: Informing the Adoption and Sustained Use of a Patient-Driven Communication Platform. Annals of Anthropological Practice 2014 Sep 02;37(2):57-74. [doi: 10.1111/napa.12029]

21. Wolcott V, Agarwal R, Nelson DA. Is Provider Secure Messaging Associated With Patient Messaging Behavior? Evidence From the US Army. J Med Internet Res 2017 Apr 06;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]

22. Baer D. Patient-Physician E-Mail Communication: The Kaiser Permanente Experience. JOP 2011 Jul;7(4):230-233. [doi: 10.1200/jop.2011.000323]

23. Keith McInnes D, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Nov 20;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]

24. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]

25. Ralston J, Cook A, Anderson M, Catz S, Fishman P, Carlson J, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control. Appl Clin Inform 2017 Dec 20;05(01):232-248. [doi: 10.4338/aci-2013-10-ra-0079]

26. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, et al. Alternatives to the Randomized Controlled Trial. Am J Public Health 2008 Aug;98(8):1359-1366. [doi: 10.2105/ajph.2007.124446]

27. Mercer SL, DeVinney BJ, Fine LJ, Green LW, Dougherty D. Study designs for effectiveness and translation research: identifying trade-offs. Am J Prev Med 2007 Aug;33(2):139-154. [doi: 10.1016/j.amepre.2007.04.005] [Medline: 17673103]

28. Rollnick S, Miller WR. What is Motivational Interviewing? Behav. Cogn. Psychother 2009 Jun 16;23(4):325-334. [doi: 10.1017/s135246580001643x]

29. Miller W, Rollnick S. Motivational Interviewing: Helping People Change, 3rd edition. New York, NY: The Guilford Press; Sep 7, 2012.

30. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Nov 01;27 Suppl 3(1):S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]

31. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. Journal of Personality and Social Psychology 1996;70(1):115-126. [doi: 10.1037/0022-3514.70.1.115]

32. Kravitz R, Bell R, Franz C. A taxonomy of requests by patients (TORP): a new system for understanding clinical negotiation in office practice. J Fam Pract 1999 Nov;48(11):872-878. [Medline: 10907624]

33. Kravitz RL, Bell RA, Franz CE, Elliott MN, Amsterdam E, Willis C, et al. Characterizing Patient Requests and Physician Responses in Office Practice. Health Serv Res 2002 Feb;37(1):215-236. [doi: 10.1111/1475-6773.00088]

34. White CB. A Content Analysis of E-mail Communication between Patients and Their Providers: Patients Get the Message. Journal of the American Medical Informatics Association 2004 Apr 02;11(4):260-267. [doi: 10.1197/jamia.m1445]

35. Shah PB. Intention-to-treat and per-protocol analysis. CMAJ 2011 Apr 05;183(6):696. [doi: 10.1503/cmaj.111-2033] [Medline: 21464181]

36. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure Messaging, Diabetes Self-management, and the Importance of Patient Autonomy: a Mixed Methods Study. J Gen Intern Med 2020 May 21:1. [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]

Abbreviations

HCCQ: Health Care Climate Questionnaire
ITT: intention to treat
MHV: My HealtheVet
MI: motivational interviewing
SAP: Supported Adoption Program
SM: secure messaging
VHA: Veterans Health Administration

Review

Classification of Depression Through Resting-State Electroencephalogram as a Novel Practice in Psychiatry: Review

Milena Čukić[1], PhD[‡]; Victoria López[2], PhD; Juan Pavón[2], PhD, Prof Dr

[1]HealthInc 3EGA, Amsterdam Health and Technology Institute, Amsterdam, Netherlands

[2]Instituto de Tecnología del Conocimiento, Institute of Knowledge Technology, Universidad Complutense Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain

[‡]Depression Detection Project

Corresponding Author:
Milena Čukić, PhD
HealthInc 3EGA
Amsterdam Health and Technology Institute
Koningin Wilhelminaplein 644
Amsterdam, 1062 KS
Netherlands
Phone: 31 615178926
Email: micu@3ega.nl

Abstract

Background: Machine learning applications in health care have increased considerably in the recent past, and this review focuses on an important application in psychiatry related to the detection of depression. Since the advent of computational psychiatry, research based on functional magnetic resonance imaging has yielded remarkable results, but these tools tend to be too expensive for everyday clinical use.

Objective: This review focuses on an affordable data-driven approach based on electroencephalographic recordings. Web-based applications via public or private cloud-based platforms would be a logical next step. We aim to compare several different approaches to the detection of depression from electroencephalographic recordings using various features and machine learning models.

Methods: To detect depression, we reviewed published detection studies based on resting-state electroencephalogram with final machine learning, and to predict therapy outcomes, we reviewed a set of interventional studies using some form of stimulation in their methodology.

Results: We reviewed 14 detection studies and 12 interventional studies published between 2008 and 2019. As direct comparison was not possible due to the large diversity of theoretical approaches and methods used, we compared them based on the steps in analysis and accuracies yielded. In addition, we compared possible drawbacks in terms of sample size, feature extraction, feature selection, classification, internal and external validation, and possible unwarranted optimism and reproducibility. In addition, we suggested desirable practices to avoid misinterpretation of results and optimism.

Conclusions: This review shows the need for larger data sets and more systematic procedures to improve the use of the solution for clinical diagnostics. Therefore, regulation of the pipeline and standard requirements for methodology used should become mandatory to increase the reliability and accuracy of the complete methodology for it to be translated to modern psychiatry.

Introduction

As the World Health Organization has warned since 2007, depression may become the most frequent cause of global disability by 2030 [1]. Only 11% to 30% of all patients diagnosed with depression reach remission within their first year of treatment [2,3]. It is possible that an individual may be diagnosed with more than one disorder (or 2 individuals showing completely different symptoms may be labeled with the same

disorder), according to the current Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) manual. Unlike many other medical specializations, psychiatry does not use objective physiological tests in its diagnostic process [4]. Many clinicians, health care providers, and researchers are aware that this diagnostic process needs improvement. Matching patients with interventions, finding specific biomarkers, and identifying various technical solutions can provide the much-needed improvement in clinical care.

Combining the knowledge and methodology used in computational neuroscience and psychiatry results in a discipline known as computational psychiatry. This field aims to determine the neurobiological underpinnings behind clusters of clinical symptoms, making it easier to adjust the treatment to patients on an individual level [5-7]. Among other studies applying this combination of approaches to resolve current issues with psychiatric diagnostics, Tokuda et al [8] offered impressive findings in their work in 2018. They combined demographic data, magnetic resonance imaging (MRI), and previous medical information on patients with applied statistical learning approaches to differentiate the 3 subtypes of depression.

Computational psychiatry may be divided into 2 approaches: theory driven and data driven. The data-driven approach typically involves some type of machine learning and appears to be much more applicable than the theory-driven approach owing to the comparably lower data collection costs. Although the most popular work published over the last period applies the data-driven approach through the use of MRI or functional MRI (fMRI) data, the drawbacks of this approach are the subject of debate among researchers. In our opinion, it would be much more appropriate to rely on electroencephalographic data, given the lower costs and higher patient accessibility. Electroencephalogram (EEG) is the oldest form of neuroimaging (1924, Hans Berger) and is noninvasive and solidly based on neurology and neuroscience. In psychiatry, it is only used to confirm the existence of epileptiform. As compared with fMRI, for example, EEG is more suitable for frequent testing owing to the lesser time required for recording and the lower price of processing. Witten and Frank [9] described data mining as "the extraction of implicit, previously unknown, and potentially useful information from the data," and at present, popular machine learning forms a part of that discipline. A typical pipeline in this framework includes recording the EEG, managing artifact removal (manually, using software, or using artifact-free epochs), linear or nonlinear electroencephalographic analysis, feature extraction, feature selection, and the application of the machine learning (both training and testing phases) method of choice. As it contains highly structured data, the EEG (a matrix of voltage values as columns/recorded from different electrode voltage and time) is highly suitable for machine learning [10].

Another research area, physiological complexity, continues to be considered novel by many medical professionals. It is based on a complex systems dynamic theory (commonly called the *chaos theory*) and is made up of vast families of distinct analysis approaches in a mathematical sense. Many researchers currently use these methods given that physiological signals are known to be nonlinear and nonstationary and generated from a highly complex system that tends to operate far from the equilibrium state. The application of a mechanistic approach (suitable for stationary signals) for the analysis of electrophysiological data, which are nonlinear, nonstationary, and noisy (3N), runs the risk of flawed interpretation. Recently, research has suggested that a mathematical link exists between the commonly applied Fourier analysis and fractal analysis [11], and the use of Fourier before the latter seems to be redundant. Klonowski [12] showed that the omnipresent classical spectral analysis in electrophysiology is based on its deeply rooted use in medicine; nonlinear analysis tends to be applied in research areas only. For a review of the varying nonlinear methodologies in detecting depression based on EEG, refer to the study by de la Torre-Luque and Bornas [13,14].

Over the past 10 years, the number of research studies using some form of machine learning on an EEG data set to detect depression or predict treatment outcomes related to the same is booming. This study aims to review the literature to offer a cross-section that is useful for determining current best practices. We have chosen to focus on the combination of physiological complexity (the application of nonlinear measures of analysis of EEG) and data-driven computational psychiatry approaches, as we believe that this combination may offer faster improvement in current clinical practices focusing on the treatment of depression.

Methods

This systematic literature review aims to find and compare published studies using nonlinear (and spectral) methods of analysis in combination with various machine learning methods for the detection of depression. Therefore, we established an inclusion criteria, as we were aware that many studies were published over the past decade. As we followed the literature for a significant amount of time, we established a start date of 2008 and an end date of May 2019.

Search Method

Given the rapid development in this research area because of faster computers, cloud utilization, and improved internet performance, we believe that this is a sufficient inclusion period. We systematically searched the Web of Science and PubMed databases on May 24, 2019, using the following combination of keywords: ("Data mining" OR "machine learning") AND ("EEG" OR "Electroencephalography") AND ("Depression" OR "MDD").

In addition, databases indexing both fields, such as Springer, Scopus, and ScienceDirect, were searched for relevant literature, including the Cornell repository.

After an original search yielding 197 papers, we reviewed all the titles and abstracts to determine which were in line with our search criteria.

Inclusion Criteria

Our eligibility criteria (eligibility testing) consisted of the following requirements: a study published between 2008 and 2019, detection of depression or predicting the outcome of treatment for depression, sample consisting of patients diagnosed

with depression (major depressive disorder [MDD]) and healthy controls (HCs), EEG data set (preferably resting-state EEG), use of fractal and nonlinear analysis as features for machine learning, and use of machine learning for detection of depression. After a primary selection phase (in which we read all the publications independently), our sample consisted of 32 publications, which was decreased to 26 based on internal discussion and comparative analysis. After reading the entire text of each publication, we decided to include 14 detection studies and 12 interventional studies. In short, we only included EEG studies that were published over the past 12 years, using task classification performed by humans with electroencephalographic signals (excluding power analyses only, nonhuman feature selection, or those with no end classification studies) that carried out a machine-based learning task aimed at detecting depression. Many studies described mobile phone apps and web-based data collection (web-based psychiatry) using machine learning, but this has already been reviewed in another work [15].

Comparisons Considered

Before conducting this systematic search, we created a list of study characteristics for comparison and to discuss the best practices and results. First, we compared the sample sizes, with only 1 intervention study being sufficiently large to analyze a sample of over 100 participants (and only 1 study consisted of only female subjects [16]). As we chose to only include EEG studies, we divided these based on resting-state EEG (employed in diagnostics) and those using any type of stimulus during the recording.

Our idea, from a nonlinear analysis perspective, is useful for analyzing resting-state records, as previous research has shown that they are the most information-rich [17]. Berman et al [18] showed that in depression, ruminative activities may only be detected in task-free and resting-state EEG recordings. Studies also varied in the number of electrodes used for recording as well as the standards used.

The next stage of comparison considered the method used for data preprocessing, some of which used standard subbands (although there is yet to be any published data or evidence that dividing EEG into subbands has any physiological significance [19]) and others used the broadband signal. Some used reductionist approaches (such as Fourier analysis or wavelet or

cosine transform) and others analyzed the raw signal. Some removed the artifact manually (probably introducing other sources of artifacts in that way), removed the artifact (automatically) with some software, or chose to analyze the epochs from artifact-free sections of recorded signal (ie, no artifact removal). Another point of discussion referred to the extent to which filtering and preprocessing were performed and whether researchers focused on any specific part of the signal's spectral content. We also compared the sampling frequency that was applied to the raw signals, an important factor for the interpretation of results. The next stage considered the type of analysis performed on previously preprocessed data and the chosen features for further machine learning.

Studies also differed in how they chose to extract or select the features.

We also noted whether internal and external cross-validation was performed (and reported) and whether the study could potentially be replicated. Finally, we compared the methods of machine learning used in each work as well as their accuracy after the testing phase and their sensitivity and specificity. Another question considered was whether the studies used receiver operating characteristic (ROC) curves to verify their accuracy. We attempted to carry out an exhaustive analysis of those publications that complied with our eligibility criteria.

Results

Diagnostic Studies

We reviewed 14 studies (classified as *detection studies*) published between 2008 and 2019. The problem with this cohort of studies is similar to that of studies trying to elucidate changes in complexity from the EEG of a patient who has depression; making a direct comparison is challenging because of the distinct methodologies employed. As we wished to draw conclusions regarding the possible practical significance of these studies and found their methodologies to be quite different in many technical aspects, we compared several basic characteristics that could potentially affect their outcomes [20]. In general, the combination of the choice of features and classification model is considered to be the most important. Nevertheless, all the studies discussed here used common processing stages, as illustrated in Figure 1.

Figure 1. Flow diagram showing common stages in the analysis of resting-state EEG in all studies with varying approaches to classification. EEG: electroencephalogram.
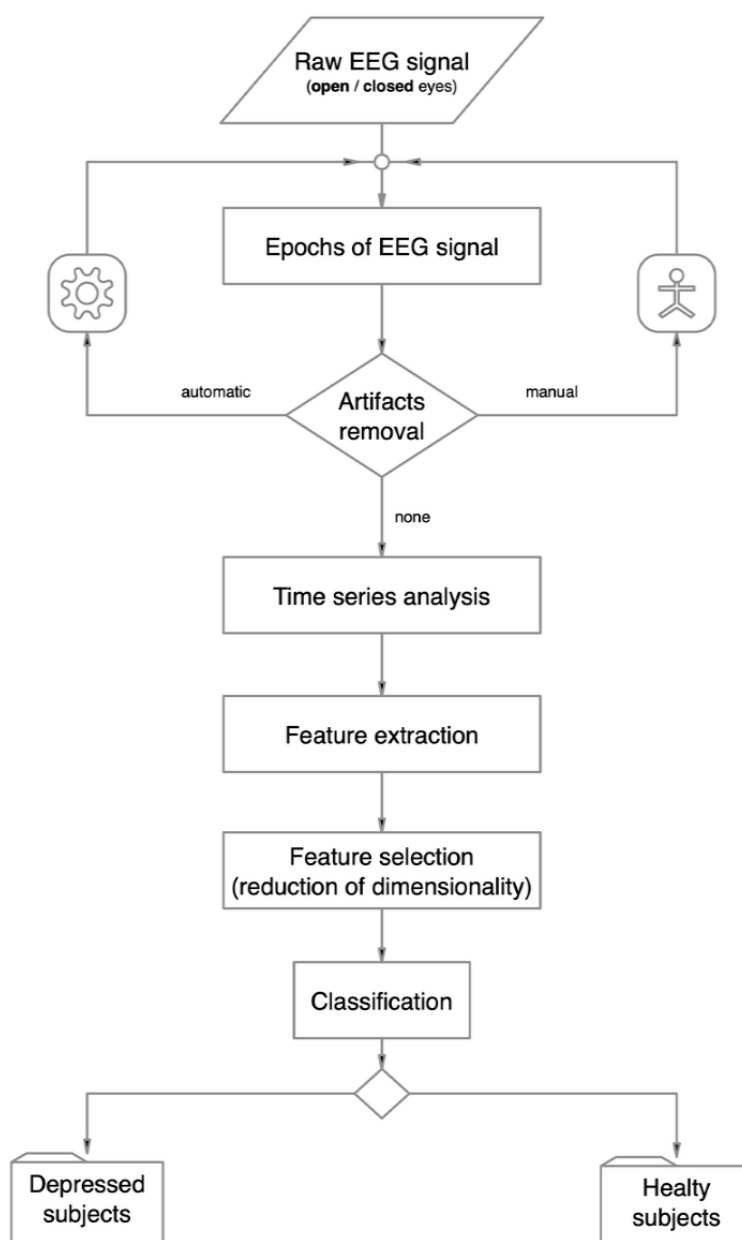
After recording the EEG (based on the previously specified method on resting-state EEG recorded with open or closed eyes, the method used to confirm the depression status, whether patients were medicated, what EEG recording standard was used, and how many electrode positions were involved), the preprocessing phase followed. Apart from standard filtering and the selection of sampling frequency, in physiological terms, the most important part involved artifact removal (manual, automatic, or no removal at all). After defining exact epochs for analysis (or better time series for further analysis), the following steps were discussed: feature extraction, feature selection (or dimensionality reduction phase), classification, validation, and the accuracy achieved in the machine learning testing phase. We also compared the conditions for study reproducibility.

Sample Size

One of the first studies using resting-state EEG to classify individuals with depression and HCs was carried out by Ahmadlou et al [21]. Their sample was quite modest (analyzing EEG recordings of 12 patients with MDD and 12 HCs). In the same year, Puthankattil and Joseph published their research [22,23] using a slightly larger sample, with 30 patients with MDD (16 males and 14 females) and 30 controls. In 2014, Hosseinifard et al [24] and Faust et al [25] published their studies based on analyses of 90 (45 MDD+45 HC) and 60 (30 MDD+30 HC) individuals. Three studies published in 2015 had similar samples: Acharya et al [26] with 30 (15 MDD+15 HC) and Bairy et al [27] with 60 (30 MDD+30 HC, probably the same or overlapping sample as in Puthankattil and Joseph [22,23] and Faust et al [25], as it has the same group of authors and the sample description was strikingly similar), whereas

Mohammadi et al [28] used a sample of 53 patients with MDD and 43 controls. A study published by Liao et al [29] also had a modest sample of 12 patients with MDD and 12 controls, whereas Mumtaz et al reported only 1 sample size used in 3 published studies with similar methodologies in 2017 [30] and 2 studies in 2018 (33 MDD+30 HC) [31,32]. Although they published their first study on spectral and fractal measures as potential markers of depression in 2013, Bachmann et al [16] added machine learning methodology to previous fractal and novel spectral (spectral asymmetry index [SASI]) analysis in their 2018 study, where their sample comprised 13 patients with MDD and 13 HCs. The study by Bachmann et al [16] is the only study that analyzed a sample consisting exclusively of female participants (which makes sense given that women present a 50% higher risk of depression). Our study [33,34] from the same year described the results of an analysis carried out on a sample of 21 patients with MDD and 20 controls. Overall, there were 362 patients with MDD and 340 age-matched HCs.

The differences between samples vary in terms of the tests used to confirm the status of the patients with MDD (DSM-IV, International Classification of Diseases, Beck Depression Scale, and Montgomery-Asberg Scale) as well as the state when recording (open or closed eyes or both). Studies also differed in terms of medication status of patients with MDD, with some being all unmedicated participants [24] with defined medication washout at 6 weeks, others being medication naïve, others stating that their patients were medicated [33], and others that did not even report on the medication status [27].

Method of Recording the EEG

Another important aspect when comparing the selected studies was how the researchers recorded the resting-state EEG, under what conditions, and using how many electrodes (concerning the standard used in the *Methods* section). Ahmadlou et al [21] analyzed a 3-min resting-state EEG, with closed eyes, from only frontal positions (using 7 out of 19 electrodes in the 10/20 standard; namely, Fp1, Fp2, Fz, F3, F4, F7, and F8), as they focused on previous findings on stable frontal asymmetry in depression. They separately analyzed the left and the right hemisphere positions and used a sampling rate of 256 Hz. Puthankattil and Joseph [22,23], Faust et al [25], and Acharya et al [26] analyzed only 4-position EEG recordings, 2 on the left hemisphere and 2 on the right hemisphere: FP1-T3 and FP2-T4; Bairy et al [27] did not report on the positions used for analyses in the *Methods* section, simply stating that only positions from the left side of the brain were taken into account. Puthankattil and Joseph [22] reported that the recording lasted for 5 min, but information on eye condition was not included; they (similar to Faust et al [25] and Acharya et al [26]) used a sampling frequency of 256 Hz. Alternately, Hosseinifard et al [24] analyzed recordings from all 19 electrodes (10/20 standard), with a sampling frequency of 1 kHz. Liao et al [29] analyzed recordings from 30 electrodes recorded for 5 min. In our research, we also analyzed all the positions on the cap and used a 1 kHz sampling rate. It is important to stress here that Bachmann et al [35], in their work from 2013, stated that they found that physiological complexity is elevated on all electrodes; therefore, they decided to use only 2 electrodes for further

analysis (and in their 2018 work [16], they focused on detection of the EEG signal from just 1 electrode). In comparison with previously mentioned studies that also used resting-state EEG, we confirmed that the number of electrodes is important, as principal component analysis (PCA) readings reveal that each electrode offers its own contribution to the results [34].

Subbands Versus Broadband

Ahmadlou et al [21] used wavelets to decompose the raw EEG signal into 5 standard subbands (gamma, beta, alpha, theta, and delta), but they also analyzed the broadband signal. Furthermore, they attempted to separately analyze the left- and right-side brain recordings but confirmed that previously significant differences disappeared when the left-right series were combined. Bachmann et al [16,35] used their own previously tested novel spectral index (SASI; based on standard subbands), and Hosseinifard et al [24] also used standard subbands. Puthankattil and Joseph [22,23] used wavelet for signal decomposition, as did Faust et al [25], whereas Acharya et al [26] (like Čukić et al [33]) used the broadband signal for analysis. Bairy et al [27] used cosine transform to decompose the signal before further analysis, although wavelets and cosine transform are also considered to be reductionist approaches, as is the Fourier approach [12]. It is not clear whether all the researchers are aware that until now no physiological significance of standard subbands of EEG use is confirmed, but their use in electrophysiology is so deeply rooted that it remains inevitable [12].

Preprocessing

Of the numerous potential options for preprocessing the signal, some common practices may be found in all the papers that were reviewed. For example, artifact removal may be performed either automatically or manually [24]. As EEG signals are always nonstationary, nonlinear, and noisy, researchers usually determine the trade-off in every aspect of the stages of analysis. For example, if the artifacts are manually removed, another type of artifact is immediately introduced into the signal. In addition, if independent component analysis (ICA) or other artifact removal techniques are used, the intrinsic signal dynamics could be changed. As stated by Goldberger et al [17], physiological signals are the richest in information when minimally preprocessed. Another option is to observe every trace and locate where the artifacts are present and subsequently select epochs for analysis from artifact-free sections (minimally changing the signal under study, as in the study by Čukić et al [33]).

We are aware that a mathematical connection exists between the Fourier transform and, for example, the Higuchi fractal dimension [11]; thus, it is clear that if one wants to calculate the fractal dimension on time series, the use of Fourier before this calculation is redundant. ICA also affects the electrophysiological signal if used for artifact removal. Perhaps those who are applying several methods of preprocessing of electrophysiological signals are unaware of some details of the consequences of signal processing.

Feature Extraction and Feature Selection

Feature extraction refers to the creation of features, such as calculating various fractal and nonlinear measures from chosen

epochs (time series) of raw signal traces. On the other hand, feature selection (or reduction of the problem dimensionality) helps to remove those features that are redundant or irrelevant. In this group of publications, different authors used different combinations of the two: Ahmadlou et al [21] used 2 different algorithms for fractal dimension calculations, whereas Higuchi and Katz used features and later attempted to compare them in terms of final accuracy. After averaging both the Higuchi fractal dimension (HFD) and the Katz fractal dimension (KFD) values, they applied the analysis of variance to assess the ability of a feature to discriminate between groups based on variations both between and within groups. Puthankattil and Joseph [22] used wavelet entropy as a feature (8-level multiresolution decomposition method of discrete wavelet transform [DWT] was used), and relative wavelet energy (RWE) analysis provided information about the signal energy distribution at different decomposition levels; 12 features were extracted for training and testing. An additional 9 features included RWE values for different frequency bands and 2 were obtained by observing the trend of the variation of the average RWE of EEG signals (RWE is higher in depression). Hosseinifard et al [24] used spectral power together with HFD, correlation dimension, and the largest Lyapunov exponent (LLE) as EEG features. Faust et al [25] used wavelet packet decomposition (WPD; Db8 wavelet) to extract appropriate subbands from the raw signal.

The results of our cross-section analysis are summarized in chronological order in Table 1.

The extracted subbands were input to calculate several entropy measures: bispectral entropy (Ph, including higher order spectra technique, from Fourier analysis), Renyi entropy, approximate entropy, and sample entropy (SampEn).

The extraction of the subband process consisted of sending the original data through a sequence of down-sampling and low-pass filters that defined the transfer function (similar to classical spectra analysis, which distorts the information content of the signal, according to Klonowski [12]). In addition, before that extraction, researchers claim that high-frequency components did not contribute relevant information (contrary to our findings [33,34]), and they were also removed. After using the Student *t* test to evaluate features, several classification algorithms were applied. Acharya et al [26] applied 15 different spectral and nonlinear measures for feature extraction: fractal dimension (HFD), LLE, SampEn, detrended fluctuation analysis (DFA),

Hurst exponent, higher order spectra features (weighted center of bispectrum, W_Bx and W_By), bispectrum phase entropy, normalized bispectral entropy and normalized bispectral squared entropies (Ent2 and Ent3), and recurrence quantification analysis parameters (determinism, entropy, laminarity [LAM], and recurrent times). These extracted features are ranked by the *t* value. There is no information as to whether they were calculated with standard EEG subbands or broadband signals (similar to the classical spectral measure or high-order spectra using Fourier analysis, they must have been computed in subbands, but this was not mentioned). After numerous trials, the authors, based on a comparison of values to formulate the Depression Diagnosis Index, decided to only consider LAM, W_By, and SampEn, without proper justification. They declared that "DDI is a unique formula that yields non-overlapping ranges for normal and depression classes." This (probably) heuristically obtained index is used here instead of the more commonly utilized classifiers [26]. Mohammadi et al [28] applied linear discriminant analysis (LDA) to map features into a new feature space (data evaluation phase) and genetic algorithm (GA) to identify the most significant features. Hosseinifard et al [24] used the leave-one-out cross-validation (LOOCV) method for the training data set. A GA was used for feature selection. The population size was established at 50 and cross-over at 80% (they also attempted PCA, but the GA significantly outperformed it). Bachmann et al [16] used the SASI spectral measure but also calculated the HFD, DFA, and Lempel-Ziv complexity (LZC) as features. In our research, we combined 2 nonlinear measures as features extracted from the signal (HFD and SampEn). Later, we decorrelated them and used PCA to reduce the dimensionality of the problem [33]. Bairy et al [27] calculated the SampEn, correlation dimension, fractal dimension, Lyapunov exponent, Hurst exponent, and DFA on DWT coefficients, and the characteristic features were ranked by *t* value. Liao et al [29] proposed a method based on scalp EEG and robust spectral spatial EEG feature extraction based on the kernel eigen-filter-bank common spatial pattern (KEFB-CSP). They first filter the multichannel EEG signals (30 electrode traces) of each subband from the original sensor space to a new space where the new signals (ie, CSPs) are optimal for the classification between patients with MDD and HCs. Finally, they implement kernel PCA to transform the vector containing the CSPs from all frequency subbands to a lower-dimensional feature vector called KEFB-CSP.

Table 1. A comparison of the previously mentioned studies comparing several characteristics, including their accuracy on the classification task.

Study	Sample (MDD[a]+HC[b])	Electrodes, frequency (Hz)	Preprocessing	Features	ML[c] models	Accuracy (%)
Ahmadlou et al, 2012 [21]	12+12	7, 256	Wavelets and spectral bands (Fourier), bootstrap	Higuchi and Katz FD[d]	Enhanced probabilistic neural networks	91.30
Puthankattil and Joseph, 2012 [22]	30 (16 M[e]+14 F[f])+30	4, 256	Wavelet, total variation filtering, multiresolution decomposition	Wavelet entropy	RWE[g], artificial feed forward networks	98.11
Hosseinifard et al, 2014 [24]	45+45	19, 1 kHz	Standard spectral bands	Power, DFA[h], Higuchi, correlation dimension, Lyapunov exponent	KNN[i], LR[j], linear discriminant	90
Faust et al, 2014 [25]	30+30	4 (2 left, 2 right), 256	Wavelet package decomposition	ApEn[k], SampEn[l], REN[m], bispectral phase entropy	PNN[n], SVM[o], DT[p], KNN, NB[q], GMM[r], Fuzzy Gueno Classifier	99.50
Bairy et al, 2015 [27]	30+30 (left brain only)	N/A[s]	Discrete cosine transform	SampEn, FD, CD[t], Hurst exp, LLE[u], DFA	DT, KNN, NB, SVM	93.80
Acharya et al, 2015 [26]	15+15	2 left, 2 right, 256	Broadband	FD, LLE, SampEn, DFA, H[v], W-Bx[w], W_By[x], EntPh[y], Ent1[z], DET[aa], ENTR[ab], LAM[ac], T2 (DDI)[ad]	SVM, KNN, NB, PNN, DT	98
Mohammadi et al, 2015 [28]	53+43	28 (10/10), 500	Standard bands/FFT[ae], LDA[af], genetic algorithm	Spectral only	DT	80
Puthankattil and Joseph, 2014 [23]	30+30	4, 256	Wavelet package decomposition	Wavelet entropy, approximate entropy	NN[ag]	98
Liao et al, 2017 [29]	12+12	30, 500	Common spatial pattern	Spectral (common spatial pattern)	KEFB-CSP[ah]	80
Mumtaz et al, 2018 [30]	34/18 F+30/9 F[ai]	19, 256	REST[aj]	Synchronization likelihood	SVM, LR, NB	87.50
Mumtaz et al, 2017 [31]	33+30	19 (EO[ak], EC[al]), 256	Fourier	Alpha interhemispheric asymmetry	LR, SVM, NB	98.40
Mumtaz et al, 2018 [32]	34+30	19, 256	10-fold cross-validation	Power, asymmetry, wavelet coefficients, Z-score	LR	94
Bachmann et al, 2018 [35]	13+13	1, 1 kHz	Fourier	HFD[am], DFA, Lempel-Ziv complexity, and SASI[an]	Logistic regression	88
Čukić et al, 2018/2020 [33,34]	26+20	19, 1 kHz	Broadband EEG[ao], 10-fold cross-validation, PCA[ap]	HFD+SampEn	MP[aq], LR, SVM (with linear and polynomial kernel), DT, RF[ar], NB	97.50

[a]MDD: major depressive disorder.

[b]HC: healthy control.

[c]ML: machine learning.

[d]FD: fractal dimension.

[e]M: male.

[f]F: female.

[g]RWE: relative wavelet energy.

[h]DFA: detrended fluctuation analysis.

[i]KNN: K-nearest neighbor.

[j]LR: linear regression.

[k]ApEn: approximate entropy.

[l]SampEn: sample entropy.

[m]REN: Renyi entropy.

[n]PNN: probabilistic neural network.

[o]SVM: support vector machine.

[p]DT: decision tree.

[q]NB: naïve Bayes.

[r]GMM: Gaussian mixture model.

[s]N/A: not applicable.

[t]CD: correlation dimension.

[u]LLE: largest Lyapunov exponent.

[v]H: Hurst exponent.

[w]W-Bx: higher order spectra features (weighted center of bispectrum [W_Bx]; Acharya et al [26]).

[x]W_By: higher order spectra features (weighted center of bispectrum [W_By]; Acharya et al [26]).

[y]EntPh: bispectrum phase entropy.

[z]Ent1: normalized bispectral entropy.

[aa]DET: determinism.

[ab]ENTR: entropy.

[ac]LAM: laminarity.

[ad]T2 (DDI): recurrent times.

[ae]FFT: fast Fourier transform.

[af]LDA: linear discriminant analysis.

[ag]NN: neural network.

[ah]KEFB-CSP: kernel eigen-filter-bank common spatial pattern.

[ai]34 depression patients (among them 18 females) and 30 healthy controls (of those 9 were female).

[aj]REST: reference electrode standardization technique.

[ak]EO: eyes opened.

[al]EC: eyes closed.

[am]HFD: Higuchi fractal dimension.

[an]SASI: spectral asymmetry index.

[ao]EEG: electroencephalogram.

[ap]PCA: principal component analysis.

[aq]MP: multilayer perceptron.

[ar]RF: random forest.

Classifiers Used and Validation

Ahmadlou et al [21] used averaged, calculated KFD and HFD values (dividing it between the left and right electrodes and averaging it) as features for enhanced probabilistic neural networks. Puthankattil and Joseph [22] used the RWE and artificial feedforward neural network, and Hosseinifard et al [24] used K-nearest neighbor (KNN), LDA, and linear regression (LR) classifiers. Two-thirds of the sample was used for the training phase and the remainder was used for the test set. Faust et al [25] used WPD (Db8 wavelet) to extract appropriate subbands from the raw signal. The extracted subbands were input for calculating the entropy measures. They used a Gaussian mixture model, decision trees (DTs), KNN, naïve Bayes classifier (NBC), probabilistic neural networks, fuzzy Sugeno classifier, and support vector machine (SVM) and 10-fold cross-validation. Acharya et al [26] used SVM with a polynomial kernel of order 3, but the validation method was not reported. Mohammadi et al [28] built predictive models

using a DT. The classifiers used in the research by Bairy et al [27] were DT, SVM, KNN, and naïve Bayes (NB). SVM employs a radial basis function. The model applied in the study by Mohammadi et al [28] revealed an average accuracy of 80% (MDD vs HC). There is no clear information regarding the verification of reliability of their high accuracy or internal and external validation (in terms of good generalization). Bachmann et al [16] used features for classification via logistic regression with LOOCV. As evident that characterization of the resting-state EEG with nonlinear measures leads to very accurate classification, we applied the 7 most popular classifiers in our research: multilayer perceptron, LR, SVM with a linear and polynomial kernel, DT, random forest, and NBC, discriminating EEG between HC subjects and patients diagnosed with depression [33]; 10-fold cross-validation was used in that work.

Classification Accuracy

Ahmadlou et al [21] found that MDD and non-MDD are more separable in the beta band based on HFD (contrary to previous

belief that differentiation is best in the alpha band) and that HFD in both beta and gamma bands is higher in patients with MDD than in healthy participants. This implied a higher complexity of signal recorded from the frontal cortices (according to their data, the left frontal lobe is more affected). On the basis of HFD (which performed better than KFD), they obtained a high accuracy of 91.3%. Puthankattil and Joseph [22] obtained artificial neural network performance with an accuracy of 98.11% (normal and depression signals). Sensitivity was 98.7%, selectivity was 97.5%, and specificity was 97.5%. In the study by Hosseinifard et al [24], classification accuracy was the highest in the alpha band for LDA and LR, both of which reached 73.3% (the worst was KNN in the delta and beta bands and LDA in the delta band with 66.6%). The highest accuracy in the experiment was obtained using the LR and LDA classifiers. The accuracy of all classifiers increased when the signal was characterized by nonlinear features, not classical power (LR reached 90% with the correlation dimension). The conclusion was that "nonlinear features give much better results in the classification of depressed patients and normal subjects," as opposed to the classical features. It was also concluded that patients with depression and controls differ more in the alpha band than in the other bands, especially in the left hemisphere [24]. Faust et al [25] applied 10-fold stratified cross-validation. The accuracy was 99.5%, sensitivity was 99.2%, and specificity was 99.7%. Contrary to Hosseinifard et al [24], they claim that the EEG signals from the right part of the brain are better at discriminating individuals with depression. In the study by Acharya et al [26], features are ranked by the t value and are fed to classifiers one by one, obtaining an accuracy of >98%, sensitivity of >97%, and specificity of >98.5%. This best result is reportedly obtained through the use of SVM with a polynomial kernel of order 3 (for both left and right hemispheres; they used averaged values for the left and right hemispheres), although SVM was discarded in previous papers by the same authors. The text is ambiguous when stating "features are fed to SVM classifier" and the following sentence "SVM classifier yielded the highest classification performance with the average accuracy..." [26]. In the study by Faust et al [25], the accuracy was 99.5%, sensitivity was 99.2%, and specificity was 99.7%. Unlike Hosseinifard et al [24], they claim that EEG signals from the right part of the brain better discriminate individuals with depression. Bairy et al [27] reported an accuracy of 93.8%, sensitivity of 92%, and specificity of 95.9%. It is impossible to determine whether internal or external validation was performed or determine the details, for example, the method used to calculate the fractal dimension (that description was not reported, limiting comparisons with our work on HFD). Thus, this study, which claims such high accuracy, has limited reproducibility. Liao et al [29] achieved 80% accuracy using KEFB-CSP, and Mumtaz et al [30] reported an SVM classification accuracy of 98%, LR classification accuracy of 91.7%, and NB classification accuracy of 93.6%. Bachmann et al [35] reached a maximum accuracy of 85% with HFD and DFA and also with HFD and LZC and, for only 1 nonlinear measure, a maximum accuracy of 77%. The average accuracy among classifiers reported by Čukić et al [33] ranged from 90.24% to 97.56%. Among the 2 measures, SampEn demonstrated better performance. When compared with the previously mentioned studies that also used resting-state EEG, it was possible to confirm that the number of electrodes is an important factor, as PCA readings demonstrate that every electrode offers its own contribution to the results [33,34].

In conclusion, we cannot state that all the mentioned studies provide sufficient information for replication, as it is clearly not the case with Bairy et al [27], who did not mention the method used to calculate the fractal dimension and the algorithm. Others concentrated on classification improvement but did not implement all the measures necessary to reach unwarranted optimism in their results. Finally, all the studies (including ours, although as declared, it was a pilot study) had very modest sample sizes, affecting the model's generalizability. A summary of comparisons of analysis of signals in the literature has been illustrated in Table 2, and a summary of comparisons with regard to the classifications applied is provided in Table 3.

Table 2. Summary of the abovementioned comparisons of analysis of signals in the literature.

Analysis of signal	Number of electrodes	Subbands	Filtering	Method of analysis	Feature extraction
Common	1, 3, or 7 (prefrontal)	Standard subbands	Preprocessing on site	Fourier and its derivatives	t test or ANOVA[a]
Recommended	19+ (all electrodes)	Broadband	Minimal preprocessing	Fractal and nonlinear	PCA[b] or GA[c]

[a]ANOVA: analysis of variance.

[b]PCA: principal component analysis.

[c]GA: genetic algorithm.

Table 3. Summary of the abovementioned comparisons with regard to the classifications applied.

Classification	Sample size	Data collection	Feature selection	Validation	Model	Accuracy
Common	12-40	1 site	Spectral analysis	Often missing	SVM[a]	Typically >95% or 99%
Recommended	>50-100	Multiple sites/collaborative (possible extraction from MRI[b] sets)	Nonlinear analysis	Internal plus external validation on unseen data	LASO[c], embedded regularization	ROC[d] curve application/more realistic results

[a]SVM: support vector machine.

[b]MRI: magnetic resonance imaging.

[c]LASO: the name of the algorithm; a type of linear regression that uses shrinkage.

[d]ROC: receiver operating characteristic.

Interventional EEG Studies

Studies have also been published during the same time interval (2008-2019) based on EEG registration, but unlike the previously mentioned work, they opted to use a stimulus (so, not resting-state EEG), a sound stimulation, or evoked response potentials (ERPs). Therefore, we briefly discuss their results. Kalatzis et al [36] published the first study on the SVM-based classification system to discriminate depression using the P600 component of ERP signals. EEG was recorded on 15 electrodes, and the sample consisted of 25 patients and an equal number of HCs. The outcomes of SVM classification were selected by the majority vote engine. Classification accuracy was reportedly 94% when using all leads and between 92% and 80% when using only the right or left electrode positions for classification. They concluded that their findings support the hypothesis that depression is associated with the dysfunction of right hemisphere mechanisms mediating the processing of information that assigns a specific response to a particular stimulus. Lee et al [37] attempted to predict the treatment response of MDD. Their study was designed to verify whether the connectivity strength of resting-state EEG could be a potential biomarker (ROC curve was 0.6 to 0.8) used to answer this question. They concluded that "...the stronger the connectivity strength, the poorer the treatment response." The experiment also suggested that frontotemporal connectivity strength could be a potential biomarker to differentiate between responders and slow responders or nonresponders in MDD. We attempted to compare our results, but their sampling frequency was as low as 100 Hz, making this comparison difficult. In a 2011 study, Cavanagh et al [38] analyzed EEG recordings from 21 medication-free patients with MDD and 24 HCs when performing a probabilistic reinforcement learning task. They measured the EEG response to error feedback, which may demonstrate selective alteration of avoidance learning, which is important in MDD. Khodayari-Rostamabad et al [39,40] probed machine learning methodology as a prediction model for a successful outcome of SSRI medication in MDD based on resting-state EEG recorded before the treatment. The sample consisted of 22 participants (11 males and 11 females). For the experiment, only 16 electrodes were used (10/20 standard) in open and closed eyes, recording for 6.5 min and combining sections into 6 files per person. The Welch model analysis yielded various spectral measures but mentioned *only as candidate features* because they did not wish to state which feature would have

predictive power in advance. After selecting the features extracted from the EEG, the authors included them in the factor analysis model, whose output is the predicted response in the form of a likelihood value; the leave-one-out randomized permutation cross-validation procedure was used for validation. For visualization (and reduction of dimensionality), they used kernelized PCA. The authors did not perform assessment on unseen samples, and they did not compare the features with HCs, relying solely on spectral measures of their modest sample. They reported an overall prediction accuracy of 87.9%.

A study from 2014 attempted to predict the depression treatment response [41]. The authors claimed that no difference exists between MDD and HC in nonlinear EEG measures (using LZC), but they somehow came to the conclusion that nonlinear measures add value to their research. They claim that theirs is the first study to use nonlinear metrics to predict the outcome of depression treatment (repetitive transcranial magnetic stimulation [rTMS] in their case). According to their reported method, the potential cause could be the focus on only one specific band and not on the analysis of broadband signals. Many subsequent researchers (and previous ones) managed to find significant differences through the use of nonlinear measures for this type of detection task [16,21,24,28,30,33]. They also claimed that they were *the first* to use complexity measures in this task. Nandrino and Pezard [41] used this approach in the analysis of EEG in depression in 1994, as did several other research groups. Bachmann et al [16] applied the same methodology (LZC) and demonstrated significant differentiation between patients and controls. Mumtaz et al [30,31] used spectral measures in several papers but found a useful difference in predicting treatment outcome in depression.

Similar to Shahaf et al [42], Etkin et al [43] applied machine learning in the task of predicting medication therapy outcomes in MDD through cognitive testing. They used pattern classification with cross-validation to determine individual patient-level composite predictive biomarkers of antidepressant outcome based on test performance and obtained 91% accuracy.

Erguzel et al [44] tested their optimized classification methods on 147 participants with MDD treated with rTMS. They tested the performance of a GA and a back-propagation neural network; they were evaluated using 6-channel pre-rTMS EEG patterns of theta and delta frequency bands. Using the reduced feature set, they obtained an increase of 0.904 in the ROC curve

(area under the curve). Zhang et al [45] explored neural complexity in patients with poststroke depression in a resting-state EEG study. Their sample consisted of 21 poststroke patients with depression (PSD), 22 patients with ischemic stroke but no depression (PSND), and 15 HCs. A total of 16 electrodes were used for recording resting-state EEG. LZC was used to assess changes in complexity from EEG. PSD (depressed) presented lower neural complexity compared with PSND (nondepressed) and control subjects for the entire brain region. LZC parameters used for the recognition of PSD possessed >85% specificity, sensitivity, and accuracy, suggesting the feasibility of LZC as a screening indicator for PSD. In addition, there were 2 antidepressive treatment nonresponse prediction studies by Shahaf et al [45] and Al-Kaysi et al [46]. Shahaf et al [45] developed a new electrophysiological attention-associated marker from a single channel (2 electrodes: Fpz and 1 earlobe) using 1-min samples with auditory oddball stimuli that was capable of detecting treatment-resistant depression (26 patients and 10 controls). Al-Kaysi et al [46] aimed to predict the transcranial direct current stimulation (tDCS) treatment outcome of patients with MDD using automated EEG classification. They accurately predicted 8 out of 10 participants when using FC4-F8 (with an accuracy of 76%) and 10 out of 10 participants when using CPz-CP2 (92% accuracy). This finding demonstrates the feasibility of using machine learning to identify patients responsive to the tDCS treatment. Cai et al [47] used only 3 electrodes on prefrontal positions to record the signal when stimulating their participants with a sound. They claim that owing to the small number of electrodes that can be easily positioned, their method has excellent potential for use in clinics. They used an electrophysiological database consisting of 92 patients with depression and 121 HCs; resting-state EEG was recorded using sound stimulation (pervasive prefrontal lobe electrodes were used on the Fp1, Fp2, and Fpz positions). After denoising (finite impulse response filter), they combined the Kalman derivative formula and the discrete wavelet transformation and an adaptive predictor filter; a total of 270 linear and nonlinear features were extracted (it is not clear what they were). Feature selection was minimal-redundancy-maximal-relevance, which reduced the dimensionality of the feature space. Four classification methods were applied: SVM, KNN, DTs, and artificial neural networks. For evaluation, they used 10-fold cross-validation. KNN presented the highest accuracy at 79.27%. Jaworska et al [48,49] published 2 papers. In the first study [48], they examined a variation of pretreatment EEG to predict depression treatment success, and in the second work [49], they performed a 12-week machine learning study to predict the outcome of pharmacology treatments in 51 patients with MDD. They used both electrophysiological and demographic data (including the Montgomery-Asberg Depression Rating Scale scores before and after treatment) as well as source-localized current density and random forest for classification, with 78% to 88% accuracy depending on model complexity. They also used kernel PCA to reduce and map important features. Similar to the abovementioned research, this lays the groundwork for studies on personalized, *biomarker*-based treatment approaches. For this group of studies, it is clear that methodology comparisons are challenging, but they are part of the same effort to show that

not only detection but also monitoring and predicting the pace of recovery, or output of the treatment (sometimes called *responders* detection), is possible. The problem with both detection and interventional studies tends to be the modest sample sizes and almost complete absence of an external validation process (on previously unseen data, from an independent sample), which puts their high reported accuracies into question.

Discussion

Most of the publications included in our review presented high accuracy in classifying individuals with depression and healthy participants based on their resting-state EEG, although they utilized various combinations of features and machine learning models. Although direct comparison is challenging, the common denominator for all presented studies can be summarized as a comparison of the methodological steps that are inevitable in this kind of research, in which certain features, previously found to be characteristic for depression, were used to feed classifiers of their choosing.

Several approaches may be used to examine the changes in the complexity of the EEG characteristic of depression. Researchers have reached a consensus that depression is characterized by high EEG complexity, compared with healthy peers [13]. Changes in functional connectivity characteristic of depression are demonstrated in the current literature, whether using fMRI, fractional anisotropy [50,51], or graph theory analysis based on EEG signals [52]. It is possible that decreased functional connectivity may be reflected by increased excitability of the cortex; thus, a difference in EEG between people diagnosed with depression and HCs could be detected [16,21,24,25]. The most important conclusion of the review by de la Torre-Luque and Bornas [13] was that "EEG dynamics for depressive patients appear more random than the dynamics of healthy non-depressed individuals." It is also accepted that the use of more than one nonlinear measure should be standard as different measures detect unique features of the EEG signals, "revealing information which other measures were unable to detect" [52,53].

The classification of patients diagnosed with depression and HCs can be considered as a first step in exploring the potential for prediction. Differentiation between episode and remission is also possible [54]. The prediction of clinical outcomes or relapses (eg, after incomplete remission or relapse in recurrent depression) would be of great clinical significance. However, there are several challenges, both methodological and statistical, to the development of a model to predict a specific clinical outcome for previously unseen individuals. A group of authors elucidated some of the risks, pitfalls, and recommended techniques to improve model reliability and validity in future research [4,7,55-58]. The authors declared that neuroimaging researchers who begin to develop such predictive models are typically unaware of some of the required considerations to accurately assess model performance and avoid inflated predictions (so-called *unwarranted optimism*) [4,55,56,59]. The common characteristics of this type of research are as follows: classification accuracy is typically 80% to 90% overall, the

sample size tends to be small to modest, and samples are usually gathered from a single site. SVM and its variants are very popular, but the use of embedded regularization frameworks is recommended, at least with the absolute shrinkage and selection operator [7]. LOOCV and k-fold cross-validation are also popular procedures for validation (for model evaluation), and model generalization capability is typically untested on independent samples [7]. The rarely employed Vapnik-Chevronenkis dimension [60] should be of standard use for model evaluation or reduction. A lack of external validation is common in most current studies. From a methodological point of view, those problems must be resolved. For example, the problem of generalization. Generalization is the ability of a model that was trained in one data set to predict patterns in another unseen data set. When testing generalizability, we are examining whether a classification is effective in an independent (not shown to the previous algorithm) population. When developing a model, one does not wish to train the classifier on a general sample characteristic; for example, if using nonlinear measures, they may differ because some measures change with age [17] or may be characteristic of a certain gender [61,62]. Some authors refer to these as *nuisance variables* because the algorithm learns to recognize that particular data set with all of its characteristics. Overfitting is common and consequently the treatment of nuisance variables. Overfitting takes place when "a developed model perfectly describes the overall aspects of the training data (including all underlying relationships and associated noise), resulting in fitting error to asymptotically become zero" [7]. Thus, the model will be unable to predict what we want on unseen (test) data. The sample size is usually small to modest (typically <100; in the study by Chekroud et al [56], for example, it is >4000 because of the use of a collaborative data set). Hence, balancing the model's complexity against the sample size is essential for improving the prediction accuracy for unseen (test) data [7]. How can this goal be achieved? By collecting more data. The collection of other more expensive neuroimaging data would be a potential solution to establish a standard set-up and start collaborative projects, as a single site is usually not sufficient to ensure a large sample size. In EEG collection, a desired model could be one that is made up of large collaborative projects such as RDoC, STAR*D, and IMAGEN. In addition, corecording with fMRI and magneto encephalography may be a solution [57]. Another option could be the use of wireless EEG caps. Although present wireless EEG caps are still quite expensive (Epoch, ENOBIO Neuroelectrics, and iMotions, to mention just a few), they can be used for research in the environment without restraining the patient or even to monitor individuals recovering from severe episodes. If wireless EEG recorders would become accessible soon, early detection and timely intervention will most likely prevail rapidly. In frameworks such as the National Institute of Mental Health Research Domain Criteria and European Roadmap for Mental Health Research, which aim to discover stratifications based on biological markers that cut across current classifications [58], this should be possible. Through large collaborative efforts, the conditions may be met to extract genuinely reliable models for clearly defined neuromarkers for future clinical use [57]. Large-scale imaging campaigns and the collection of general population data are essential conditions for the transfer of these research findings to clinics. By permitting regular medical checkup data to become a part of such organized collaborative efforts, patients would also contribute to the improvement of this precise diagnostic in the near future. According to Kraiij [63], the 4P concept for health care improvement stands for prediction, prevention, personalization, and participation. It has been suggested that health care focuses too much on disease treatment and not enough on its prevention. It has also been observed [58] that treatment and diagnosis tend to be based on population averages. In some cases, treatment has negative effects. Therefore, there is much room for improvement (and for the other 3 Ps aside, personalization). Data collection, analysis, and sharing play an important role in the improvement of health care. The first project to implement the 4Ps is the SWELL (Smart Reasoning for Well-being at Home and at Work) project, part of the Dutch national ICT program, COMMIT (between 2011 and 2016 in the Netherlands, Leiden University). There is also an option of testimonial data sharing that is already official, eg, in Austria.

Whelan and Garavan [55] addressed overfitting and many other methodological issues. They revealed how regression models may incorrectly appear to be predictive. They also described methods for quantifying and improving model reliability and validity. The authors conclude that "...perhaps counterintuitively to those who deal primarily with a general linear model, optimism increases as a function of the decreasing number of participants and the increasing number of predictor variables in the model (the model appears better as sample size decreases)" [55].

Although it has been shown that small sample sizes and a lack of external validation lead to unwarranted optimism, most published research does not embrace these principles as standard practice [64]. Collecting additional data may resolve this issue. The theory of data mining is clear; all models work best on larger samples. The repository may be used to test a developed model on an unseen cohort. We learned that statistics needs "to stop making fools of ourselves" [65]. Data mining is the art of finding meaning from supposedly meaningless data. Peduzzi et al [66] showed the optimal number of events per variable in logistic regression analysis.

A minimum rate of 10 cases per predictor is common [64], although it is not a universal recommendation [67]. Optimism may also be avoided with the introduction of the regularization term [68]. In addition, using previous information to constrain model complexity relying on Bayesian approaches is recommended. Bootstrapping [69] is another useful method, as is cross-validation [59]. Cross-validation tests the model's ability to generalize and involves separating the data into subsets. Both Kohavi [70] and Ng [64] described this technique. In addition, an effective and efficient 10-fold cross-validation, Elastic Net, is useful for optimizing parameters. Ng [64] stated that "...optimism becomes unreliable as the probability of overfitting to the test data increases with multiple comparisons." One can use several functions available in MATLAB (MathWorks) such as lassoglm, bootstrap for bootstrap sampling, or several functions for Bayesian analysis or the function *crossvalind* for testing sets and cross-validation.

In conclusion, when discussing the importance of maintaining completely separate training and test subsets, Whelan and Garavan [55] stated the following: "any cross-contamination will result in optimism." We could not agree more. Additional research is necessary to reframe nosology in psychiatry and to help support the patient's journey to remission. We hope that many people will benefit from the cloud-based services provided by already digitized health care institutions.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization. 2017. URL: http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf [accessed 2020-10-07]

2. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-1917. [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]

3. Cipriani A, Furukawa T, Salanti G, Geddes J, Higgins J, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009 Feb;373(9665):746-758. [doi: 10.1016/S0140-6736(09)60046-5] [Medline: 758]

4. Gillan CM, Whelan R. What big data can do for treatment in psychiatry. Curr Opin Behav Sci 2017 Dec;18:34-42. [doi: 10.1016/j.cobeha.2017.07.003] [Medline: 53198151]

5. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci 2012 Jan;16(1):72-80 [FREE Full text] [doi: 10.1016/j.tics.2011.11.018] [Medline: 22177032]

6. Wang X, Krystal J. Computational psychiatry. Neuron 2014 Nov 05;84(3):638-654 [FREE Full text] [doi: 10.1016/j.neuron.2014.10.018] [Medline: 25442941]

7. Yahata N, Kasai K, Kawato M. Computational neuroscience approach to biomarkers and treatments for mental disorders. Psychiatry Clin Neurosci 2017 Apr;71(4):215-237 [FREE Full text] [doi: 10.1111/pcn.12502] [Medline: 28032396]

8. Tokuda T, Yoshimoto J, Shimizu Y, Okada G, Takamura M, Okamoto Y, et al. Identification of depression subtypes and relevant brain regions using a data-driven approach. Sci Rep 2018 Sep 20;8(1):14082 [FREE Full text] [doi: 10.1038/s41598-018-32521-z] [Medline: 30237567]

9. Witten IH, Frank E, Hall MA. Measuring the social interactions of people with traumatic brain injury and their communication partners: The adapted Kagan scales. In: DATA MINING Practical Machine Learning Tools and Techniques. New York, USA: Elsevier; 2005.

10. Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng 2019 Jun;16(3):031001. [doi: 10.1088/1741-2552/ab0ab5] [Medline: 30808014]

11. Kalauzi A, Bojić T, Vuckovic A. Modeling the relationship between Higuchi's fractal dimension and Fourier spectra of physiological signals. Med Biol Eng Comput 2012 Jul;50(7):689-699. [doi: 10.1007/s11517-012-0913-9] [Medline: 22588703]

12. Klonowski W. From conformons to human brains: an informal overview of nonlinear dynamics and its applications in biomedicine. Nonlinear Biomed Phys 2007 Jul 5;1(1):5 [FREE Full text] [doi: 10.1186/1753-4631-1-5] [Medline: 17908344]

13. de la Torre-Luque A, Bornas X. Complexity and irregularity in the brain oscillations of depressive patients: a systematic review. Neuropsychiatry 2017;7(5). [doi: 10.4172/neuropsychiatry.1000238]

14. Nandrino J, Pezard L, Martinerie J, el Massioui F, Renault B, Jouvent R, et al. Decrease of complexity in EEG as a symptom of depression. Neuroreport 1994 Jan 12;5(4):528-530. [doi: 10.1097/00001756-199401120-00042] [Medline: 8003689]

15. Llamocca PP, Junestrand M, Cukic M, Urgeles D, Lopez V. Data Source Analysis in Mood Disorder Research. In: XVIII Conference of the Spanish Association of Artificial Intelligence. 2018 Presented at: CAEPIA; October 23-27, 2018; Granada, Spain.

16. Bachmann M, Päeske L, Kalev K, Aarma K, Lehtmets A, Ööpik P, et al. Methods for classifying depression in single channel EEG using linear and nonlinear signal analysis. Comput Methods Programs Biomed 2018 Mar;155:11-17. [doi: 10.1016/j.cmpb.2017.11.023] [Medline: 29512491]

17. Goldberger AL, Peng C, Lipsitz LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging 2002;23(1):23-26. [doi: 10.1016/s0197-4580(01)00266-4] [Medline: 11755014]

XSL•FO

RenderX

18.   Berman MG, Peltier S, Nee DE, Kross E, Deldin PJ, Jonides J. Depression, rumination and the default network. Soc Cogn Affect Neurosci 2011 Oct;6(5):548-555 [FREE Full text] [doi: 10.1093/scan/nsq080] [Medline: 20855296]

19.   Başar E, Güntekin B, Atagün I, Turp Gölbaşı B, Tülay E, Ozerdem A. Brain's alpha activity is highly reduced in euthymic bipolar disorder patients. Cogn Neurodyn 2012 Feb;6(1):11-20 [FREE Full text] [doi: 10.1007/s11571-011-9172-y] [Medline: 23372616]

20.   Flach P. Machine Learning: the Art and Science of Algorithms That Make Sense of Data. Cambridge, USA: Cambridge University Press; 2012.

21.   Ahmadlou M, Adeli H, Adeli A. Fractality analysis of frontal brain in major depressive disorder. Int J Psychophysiol 2012 Aug;85(2):206-211. [doi: 10.1016/j.ijpsycho.2012.05.001] [Medline: 22580188]

22.   Puthankattil SD, Joseph PK. Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy. J Mech Med Biol 2012 Oct 23;12(04):1240019. [doi: 10.1142/S0219519412400192]

23.   Puthankattil SD, Joseph P. Analysis of EEG signals using wavelet entropy and approximate entropy: a case study on depression patients. Int J Medical Heal Biomed Bioeng Pharm Eng 2014;8(7):424.

24.   Hosseinifard B, Moradi MH, Rostami R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Comput Methods Programs Biomed 2013 Mar;109(3):339-345. [doi: 10.1016/j.cmpb.2012.10.008] [Medline: 23122719]

25.   Faust O, Ang PC, Puthankattil SD, Joseph PK. Depression diagnosis support system based on EEG signal entropies. J Mech Med Biol 2014 Mar 13;14(03):1450035. [doi: 10.1142/s0219519414500353]

26.   Acharya UR, Sudarshan VK, Adeli H, Santhosh J, Koh JE, Puthankatti SD, et al. A novel depression diagnosis index using nonlinear features in EEG signals. Eur Neurol 2015;74(1-2):79-83. [doi: 10.1159/000438457] [Medline: 26303033]

27.   Bairy GM, Bhat S, Eugene LW, Niranjan UC, Puthankattil SD, Joseph PK. Automated classification of depression electroencephalographic signals using discrete cosine transform and nonlinear dynamics. J Med Imaging Hlth Inform 2015 Jun 01;5(3):635-640. [doi: 10.1166/jmihi.2015.1418]

28.   Mohammadi M, Al-Azab F, Raahemi B, Richards G, Jaworska N, Smith D, et al. Data mining EEG signals in depression for their diagnostic value. BMC Med Inform Decis Mak 2015 Dec 23;15:108 [FREE Full text] [doi: 10.1186/s12911-015-0227-6] [Medline: 26699540]

29.   Liao S, Wu C, Huang H, Cheng W, Liu Y. Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. Sensors (Basel) 2017 Jun 14;17(6):1-18 [FREE Full text] [doi: 10.3390/s17061385] [Medline: 28613237]

30.   Mumtaz W, Xia L, Ali SS, Yasin MA, Hussain M, Malik AS. Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). Biomedical Signal Processing and Control 2017 Jan;31:108-115. [doi: 10.1016/j.bspc.2016.07.006]

31.   Mumtaz W, Xia L, Mohd Yasin MA, Azhar Ali SS, Malik AS. A wavelet-based technique to predict treatment outcome for major depressive disorder. PLoS One 2017;12(2):e0171409 [FREE Full text] [doi: 10.1371/journal.pone.0171409] [Medline: 28152063]

32.   Mumtaz W, Ali SS, Yasin MA, Malik AS. A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). Med Biol Eng Comput 2018 Feb;56(2):233-246. [doi: 10.1007/s11517-017-1685-z] [Medline: 28702811]

33.   Čukić M, Pokrajac D, Stokić M, Simić S, Radivojević V, Ljubisavljević M. EEG machine learning with Higuchi's fractal dimension and Sample Entropy as features for successful detection of depression. arXiv 2018:- [FREE Full text]

34.   Čukić M, Stokić M, Simić S, Pokrajac D. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. Cogn Neurodyn 2020 Aug;14(4):443-455. [doi: 10.1007/s11571-020-09581-x] [Medline: 32655709]

35.   Bachmann M, Lass J, Suhhova A, Hinrikus H. Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram. Comput Math Methods Med 2013;2013:251638 [FREE Full text] [doi: 10.1155/2013/251638] [Medline: 24232245]

36.   Kalatzis I, Piliouras N, Ventouras E, Papageorgiou C, Rabavilas A, Cavouras D. Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. Comput Methods Programs Biomed 2004 Jul;75(1):11-22. [doi: 10.1016/j.cmpb.2003.09.003] [Medline: 15158043]

37.   Lee T, Wu Y, Yu YW, Chen M, Chen T. The implication of functional connectivity strength in predicting treatment response of major depressive disorder: a resting EEG study. Psychiatry Res 2011 Dec 30;194(3):372-377. [doi: 10.1016/j.pscychresns.2011.02.009] [Medline: 22041534]

38.   Cavanagh JF, Bismark AJ, Frank MJ, Allen JJ. Larger error signals in major depression are associated with better avoidance learning. Front Psychol 2011;2:331 [FREE Full text] [doi: 10.3389/fpsyg.2011.00331] [Medline: 22084638]

39.   Khodayari-Rostamabad A, Hasey GM, Maccrimmon DJ, Reilly JP, de Bruin H. A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. Clin Neurophysiol 2010 Dec;121(12):1998-2006. [doi: 10.1016/j.clinph.2010.05.009] [Medline: 21035741]

40. Khodayari-Rostamabad A, Reilly JP, Hasey GM, de Bruin H, Maccrimmon DJ. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. Clin Neurophysiol 2013 Oct;124(10):1975-1985. [doi: 10.1016/j.clinph.2013.04.010] [Medline: 23684127]

41. Arns M, Cerquera A, Gutiérrez RM, Hasselman F, Freund JA. Non-linear EEG analyses predict non-response to rTMS treatment in major depressive disorder. Clin Neurophysiol 2014 Jul;125(7):1392-1399. [doi: 10.1016/j.clinph.2013.11.022] [Medline: 24360132]

42. Shahaf G, Yariv S, Bloch B, Nitzan U, Segev A, Reshef A, et al. A pilot study of possible easy-to-use electrophysiological index for early detection of antidepressive treatment non-response. Front Psychiatry 2017;8:128 [FREE Full text] [doi: 10.3389/fpsyt.2017.00128] [Medline: 28769825]

43. Etkin A, Patenaude B, Song YJ, Usherwood T, Rekshan W, Schatzberg AF, et al. A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. Neuropsychopharmacology 2015 May;40(6):1332-1342 [FREE Full text] [doi: 10.1038/npp.2014.333] [Medline: 25547711]

44. Erguzel TT, Ozekes S, Tan O, Gultekin S. Feature selection and classification of electroencephalographic signals: an artificial neural network and genetic algorithm based approach. Clin EEG Neurosci 2015 Oct;46(4):321-326. [doi: 10.1177/1550059414523764] [Medline: 24733718]

45. Zhang Y, Wang C, Sun C, Zhang X, Wang Y, Qi H, et al. Neural complexity in patients with poststroke depression: a resting EEG study. J Affect Disord 2015 Dec 1;188:310-318 [FREE Full text] [doi: 10.1016/j.jad.2015.09.017] [Medline: 26402253]

46. Al-Kaysi AM, Al-Ani A, Loo CK, Powell TY, Martin DM, Breakspear M, et al. Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. J Affect Disord 2017 Jan 15;208:597-603. [doi: 10.1016/j.jad.2016.10.021] [Medline: 28029427]

47. Cai H, Han J, Chen Y, Sha X, Wang Z, Hu B, et al. A pervasive approach to EEG-based depression detection. Complexity 2018;2018:1-13. [doi: 10.1155/2018/5238028]

48. Jaworska N, Wang H, Smith DM, Blier P, Knott V, Protzner AB. Pre-treatment EEG signal variability is associated with treatment success in depression. Neuroimage Clin 2018;17:368-377 [FREE Full text] [doi: 10.1016/j.nicl.2017.10.035] [Medline: 29159049]

49. Jaworska N, de la Salle S, Ibrahim M, Blier P, Knott V. Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data. Front Psychiatry 2018;9:768 [FREE Full text] [doi: 10.3389/fpsyt.2018.00768] [Medline: 30692945]

50. de Kwaasteniet B, Ruhe E, Caan M, Rive M, Olabarriaga S, Groefsema M, et al. Relation between structural and functional connectivity in major depressive disorder. Biol Psychiatry 2013 Jul 1;74(1):40-47. [doi: 10.1016/j.biopsych.2012.12.024] [Medline: 23399372]

51. Vederine F, Wessa M, Leboyer M, Houenou J. A meta-analysis of whole-brain diffusion tensor imaging studies in bipolar disorder. Prog Neuropsychopharmacol Biol Psychiatry 2011 Dec 1;35(8):1820-1826. [doi: 10.1016/j.pnpbp.2011.05.009] [Medline: 21624424]

52. Kim D, Bolbecker AR, Howell J, Rass O, Sporns O, Hetrick WP, et al. Disturbed resting state EEG synchronization in bipolar disorder: a graph-theoretic analysis. Neuroimage Clin 2013;2:414-423 [FREE Full text] [doi: 10.1016/j.nicl.2013.03.007] [Medline: 24179795]

53. Burns TF, Rajan R. Combining complexity measures of EEG data: multiplying measures reveal previously hidden information. F1000 Research 2015:- [FREE Full text] [doi: 10.12688/f1000research.6590.1]

54. Čukić M, Stokić M, Radenković S, Ljubisavljević M, Simić S, Savić D. Nonlinear analysis of EEG complexity in episode and remission phase of recurrent depression. Int J Methods Psychiatr Res 2020 Jun;29(2):e1816 [FREE Full text] [doi: 10.1002/mpr.1816] [Medline: 31820528]

55. Whelan R, Garavan H. When optimism hurts: inflated predictions in psychiatric neuroimaging. Biol Psychiatry 2014 May 1;75(9):746-748. [doi: 10.1016/j.biopsych.2013.05.014] [Medline: 23778288]

56. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/S2215-0366(15)00471-X] [Medline: 26803397]

57. Gillan C, Daw N. Taking psychiatry research online. Neuron 2016 Jul 6;91(1):19-23 [FREE Full text] [doi: 10.1016/j.neuron.2016.06.002] [Medline: 27387647]

58. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. Biol Psychiatry Cogn Neurosci Neuroimaging 2016 Sep;1(5):433-447 [FREE Full text] [doi: 10.1016/j.bpsc.2016.04.002] [Medline: 27642641]

59. Efron R, Tibshirani R. Improvements on Cross-Validation: The .632+ Bootstrap Method. J Am Stat Assoc 2020:548 [FREE Full text]

60. Vapnik VN. Statistical Learning Theory. Hoboken, New Jersey: John Wiley & Sons; 1998:42 URL: http://read.pudn.com/downloads161/ebook/733192/Statistical-Learning-Theory.pdf

61. Ahmadi K, Ahmadlou M, Rezazade M, Azad-Marzabadi E, Sajedi F. Brain activity of women is more fractal than men. Neurosci Lett 2013 Feb 22;535:7-11. [doi: 10.1016/j.neulet.2012.12.043] [Medline: 23313595]

62.   Ahmadlou M, Adeli H, Adeli A. Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men. Clin EEG Neurosci 2013 Jul;44(3):175-181. [doi: 10.1177/1550059413480504] [Medline: 23545250]

63.   Kraiij W. Innaugurational speech on the acceptance of his position as professor of Applied data Analytics at University Leiden. Universiteit Leiden. 2017 Feb 24. URL: https://www.universiteitleiden.nl/en/research/research-output/science/inaugural-lecture-data-of-value

64.   Ng AY. Preventing overfitting of cross-validation data. In: Proceedings of the 14th International Conference on Machine Learning. 1997 Presented at: ICML'97; July 23, 1997; Nashville, Tennessee URL: http://robotics.stanford.edu/~ang/papers/cv-final.pdf

65.   Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]

66.   Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [doi: 10.1016/s0895-4356(96)00236-3]

67.   Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007 Mar 15;165(6):710-718. [doi: 10.1093/aje/kwk052] [Medline: 17182981]

68.   Moons K, Donders AR, Steyerberg E, Harrell F. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. J Clin Epidemiol 2004 Dec;57(12):1262-1270. [doi: 10.1016/j.jclinepi.2004.01.020] [Medline: 15617952]

69.   Efron B, Tibshirani RJ. An introduction to the Bootstrap. London, United Kingdom: Chapman & Hall; 1993:- URL: http://www.hms.harvard.edu/bss/neuro/bornlab/nb204/statistics/bootstrap.pdf

70.   Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selecti. 1995 Presented at: International Joint Conference on Artificial Intelligence; 1995; Montreal, Quebec, Canada p. 1137-1143 URL: https://dl.acm.org/doi/10.5555/1643031.1643047

Abbreviations

DFA: detrended fluctuation analysis

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, fourth edition

DT: decision tree

DWT: discrete wavelet transform

EEG: electroencephalogram

ERP: evoked response potential

fMRI: functional magnetic resonance imaging

GA: genetic algorithm

HC: healthy control

HFD: Higuchi fractal dimension

ICA: independent component analysis

KEFB-CSP: kernel eigen-filter-bank common spatial pattern

KFD: Katz fractal dimension

KNN: K-nearest neighbors

LAM: laminarity

LDA: linear discriminant analysis

LLE: largest Lyapunov exponent

LOOCV: leave-one-out cross-validation

LR: linear regression

LZC: Lempel-Ziv complexity

MDD: major depressive disorder

MRI: magnetic resonance imaging

NB: naïve Bayes

NBC: naïve Bayes classifier

PCA: principal component analysis

PSD: poststroke patients with depression

PSND: patients with ischemic stroke but no depression

ROC: receiver operating characteristic

rTMS: repetitive transcranial magnetic stimulation

RWE: relative wavelet energy

SampEn: sample entropy

SASI: spectral asymmetry index

SVM: support vector machine

XSL•FO

RenderX

tDCS: transcranial direct current stimulation
WPD: wavelet packet decomposition

XSL•FO
RenderX

<u>Original Paper</u>

Automated Diagnosis of Various Gastrointestinal Lesions Using a Deep Learning–Based Classification and Retrieval Framework With a Large Endoscopic Database: Model Development and Validation

Muhammad Owais[1], MSc; Muhammad Arsalan[1], PhD; Tahir Mahmood[1], MSc; Jin Kyu Kang[1], BSc; Kang Ryoung Park[1], PhD

Division of Electronics and Electrical Engineering, Dongguk University, Seoul, Republic of Korea

Corresponding Author:
Kang Ryoung Park, PhD
Division of Electronics and Electrical Engineering
Dongguk University
30 Pildong-ro 1-gil, Jung-gu
Seoul, 04620
Republic of Korea
Phone: 82 10 3111 7022
Fax: 82 2 2277 8735
Email: parkgr@dgu.edu

Abstract

Background: The early diagnosis of various gastrointestinal diseases can lead to effective treatment and reduce the risk of many life-threatening conditions. Unfortunately, various small gastrointestinal lesions are undetectable during early-stage examination by medical experts. In previous studies, various deep learning–based computer-aided diagnosis tools have been used to make a significant contribution to the effective diagnosis and treatment of gastrointestinal diseases. However, most of these methods were designed to detect a limited number of gastrointestinal diseases, such as polyps, tumors, or cancers, in a specific part of the human gastrointestinal tract.

Objective: This study aimed to develop a comprehensive computer-aided diagnosis tool to assist medical experts in diagnosing various types of gastrointestinal diseases.

Methods: Our proposed framework comprises a deep learning–based classification network followed by a retrieval method. In the first step, the classification network predicts the disease type for the current medical condition. Then, the retrieval part of the framework shows the relevant cases (endoscopic images) from the previous database. These past cases help the medical expert validate the current computer prediction subjectively, which ultimately results in better diagnosis and treatment.

Results: All the experiments were performed using 2 endoscopic data sets with a total of 52,471 frames and 37 different classes. The optimal performances obtained by our proposed method in accuracy, F1 score, mean average precision, and mean average recall were 96.19%, 96.99%, 98.18%, and 95.86%, respectively. The overall performance of our proposed diagnostic framework substantially outperformed state-of-the-art methods.

Conclusions: This study provides a comprehensive computer-aided diagnosis framework for identifying various types of gastrointestinal diseases. The results show the superiority of our proposed method over various other recent methods and illustrate its potential for clinical diagnosis and treatment. Our proposed network can be applicable to other classification domains in medical imaging, such as computed tomography scans, magnetic resonance imaging, and ultrasound sequences.

XSL•FO
RenderX

Introduction

Various types of gastrointestinal (GI) disorders, such as a tumors, ulcerative colitis, irritable bowel syndrome, hemorrhoids, *Helicobacter pylori*, Crohn disease, polyps, and colorectal cancer, are among the leading causes of death [1]. In the United States, about 76,940 people died in 2016 due to different types of gastric cancers, according to the American Cancer Society [1]. Early and accurate diagnosis of severe diseases, such as polyps or tumors, using endoscopy videos is of great significance and leads to better treatment. However, the subjective diagnosis of such GI diseases is not only a tedious and time-consuming task but also requires sufficient knowledge and clinical experience. These diagnostic problems can be solved to a great extent by developing effective computer-aided diagnosis (CAD) tools that provide a fully automated way of detecting and classifying different GI diseases. CAD tools can assist medical experts in effective diagnosis and treatment during the initial stage of severe medical conditions [2-10]. Figure S1 in Multimedia Appendix 1 presents an overall workflow diagram of a CAD tool to visualize its clinical usability and significance in making a diagnostic decision. In the first step, medical professionals use a particular type of imaging modality to visualize the internal structure of body organs, such as the GI tract. After that, a CAD model analyzes the visual data (obtained in the first step) to highlight the lesions or suspicious regions. Finally, these highlighted results further assist the medical experts in making an effective diagnostic decision in a short time.

In the last few years, deep learning algorithms, especially artificial neural networks (ANNs), have significantly enhanced the performance of artificial intelligence–based CAD tools, which are used for diagnostic purposes in various medical domains [11-15]. In general, these ANN models undergo a training procedure to learn the optimal representation of the training data set [16] by using optimization algorithms, such as stochastic gradient descent [17]. In this way, a deep learning–based trained model that contains the optimal representation of the training data set in trainable parameters is obtained. In clinical practice, this model can analyze newly acquired endoscopic videos or images by using previous knowledge of trainable parameters. Various types of ANN models have been proposed in the image analysis domain. Among these models, convolutional neural networks (CNNs) [16] have gained special attention due to their superior performance in various image recognition–based applications, including in medical fields. The convolutional layers are considered the key part of a CNN model and contain trainable filters of different depths and sizes. These filters are trained during the training procedure by extracting complex hidden patterns (also known as deep features) from the training data set.

Over the past few years, considerable contributions have been made by robust and efficient CAD tools in the endoscopy domain. However, most of these methods are designed to detect specific types of GI diseases, such as polyps, ulcers, tumors, or cancer, using handcrafted or deep features–based approaches. Before the advent of deep features–based methods, most studies used handcrafted features such as color and texture information to perform the automated detection and classification of particular types of GI disease [18-24]. In recent years, various deep learning–based CAD tools have been proposed for endoscopic video and image analysis [2-10]. Such deep learning–based CAD tools are capable of performing the classification and detection of different GI abnormalities in a more precise and accurate way than the previous handcrafted features–based methods. However, most of the previous deep learning–based CAD methods used only the spatial information for the automatic diagnosis of GI diseases, which reduced the overall diagnostic performance. The internal structure of the human GI tract is captured as a moving sequence (video) with respect to time during an endoscopy procedure. Therefore, an endoscopic video encompasses both spatial and temporal information. In a video, the temporal information exists in the sequence of consecutive frames and provides essential information. Therefore, it is possible to use both spatial and temporal information in developing a high-performance CAD tool with good diagnostic capability. A comprehensive analysis of existing studies [2-10,18-24] in comparison with our proposed method can be found in Multimedia Appendix 1.

The motivation behind of this study was the development of a comprehensive CAD framework that would be able to recognize a wide range of GI diseases simultaneously rather than multiple CAD tools used one by one to detect an anonymous GI disease. To accomplish this task, we considered a total of 37 different classes (including both diseased and normal cases) related to the human GI tract, which is significantly more than most recent studies. Another motivation was the included cross-validation mechanism in the proposed CAD tool that provides visual information about its diagnostic decision. Such additional information can assist medical experts in validating the computer decision interactively. Therefore, in this research work, we use the strength of recent AI techniques in the endoscopy domain and propose a high-performance classification and retrieval framework for multiple GI diseases using endoscopic videos. Mainly, the overall pipeline of the proposed classification network is composed of a densely connected convolutional network (DenseNet), our defined long short-term memory (LSTM) network using LSTM cells, principal component analysis (PCA), and the k-nearest neighbors (KNN) algorithm. Experimental results demonstrate the superiority of the proposed CAD framework in comparison with various state-of-the-art methods. This study provides five main contributions.

First, this is the first spatiotemporal feature–based CAD framework based on the integrated DenseNet and LSTM followed by a PCA-based KNN classifier for the effective diagnosis of various GI diseases.

Second, with the addition of PCA, our method reduces the feature dimension up to 95%, with the gain of an average accuracy of 3.62% in comparison with previous work [10].

Third, we include the retrieval framework after the classification network to validate the CAD decision subjectively.

Fourth, in our framework, the number of successive frames to be classified can be variable rather than using the fixed-length sequence.

Fifth, we have made our trained model publicly available through Dongguk University [25], along with the information regarding training and testing data splitting.

Methods

Study Design

In this section, a detailed description of our proposed method is formulated in sequential order. An overall description of a class prediction–based video or image retrieval system using our proposed spatiotemporal feature–based classification network is given, followed by a detailed explanation of the proposed classification network, which includes spatiotemporal feature extraction and classification stages.

An Overview of the Proposed Approach

In general image or video classification and retrieval frameworks, the key element is the optimal representation of visual information or features. The optimal features are further used in retrieving relevant information from the image or video database based on feature-matching mechanisms, such as the minimum Euclidian distance. Thus, the overall performance of such a system is directly related to the methods that perform the optimal feature extraction task. Recently, deep learning–based feature extraction methods have shown the best performance in various image- and video-processing domains. Therefore, in our proposed framework, the strength of such deep learning–based algorithms was used to obtain high classification and high retrieval performance in endoscopy. A comprehensive flow diagram of our proposed framework is shown in Figure 1.

First, deep feature–based spatial and temporal information was extracted from the given input endoscopy video of n successive frames by using a cascaded DenseNet and LSTM-based network consecutively. In this way, a collective spatiotemporal feature vector was obtained for the given endoscopy video, which was further used at the classification stage (after applying the PCA [26]) to predict the class label. Second, the predicted class label was used to select the set of relevant class features from the feature database, and these were further used at the feature matching stage. Next, the extracted spatiotemporal feature vector (obtained in the first step from the input query sequence) was matched with the set of selected features (obtained in the second step), and retrieval information (ie, frame ID) was obtained based on the best-matching results. Finally, class prediction–based retrieval was done by selecting the best-matched cases from the entire database based on retrieval information.

Figure 1. Comprehensive flow diagram of the proposed classification and retrieval framework. The red dotted box highlights our major contributions in this proposed retrieval framework.
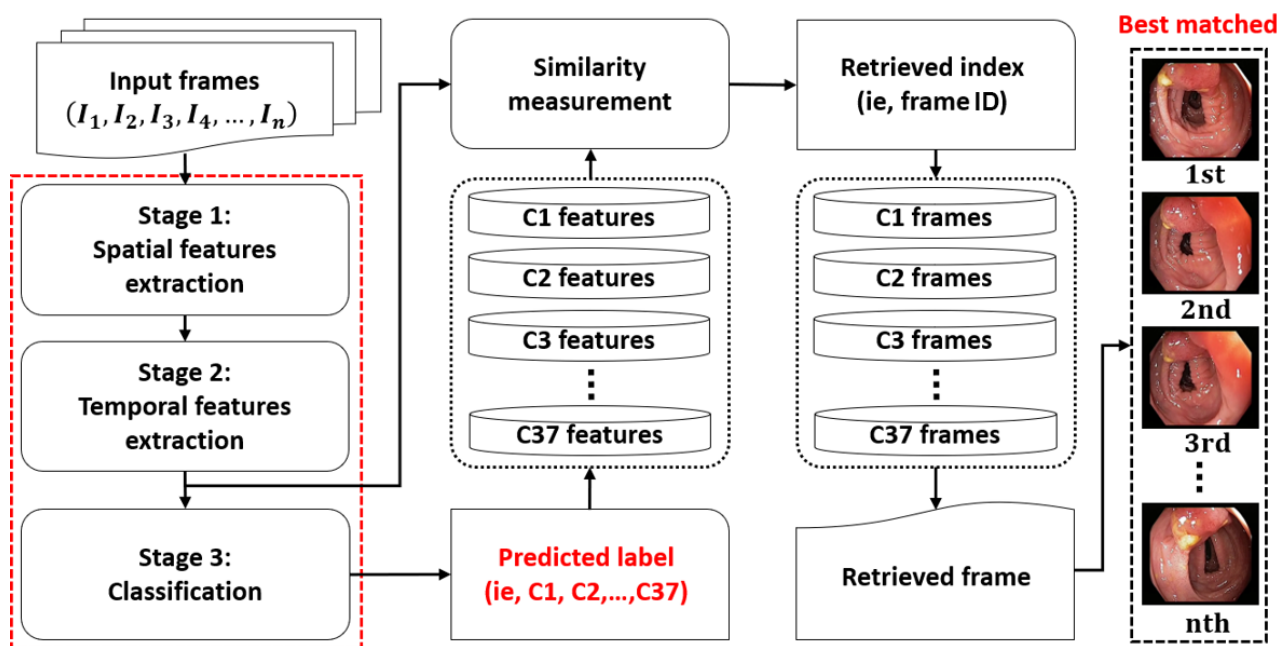
The Structure of Our Proposed Spatiotemporal Feature–Based Classification Model

Our proposed spatiotemporal feature–based classification model is composed of a DenseNet followed by the LSTM model to extract spatial and temporal features, respectively. Figure 2 presents an overall block diagram of our proposed spatiotemporal feature–based classification model. For better understanding, the complete structure of our model is divided into three main stages. In the first stage, each frame of the given input video of n consecutive frames (ie, I_1, I_2, I_3…, I_n) is processed by the DenseNet to extract the spatial features. Here, the factor n presents the length of the input frames, which control the span of temporal features with respect to time. The total time span of n consecutive frames can be calculated by multiplying n with the frame rate, which is 30 (ie, 30 frames per second) in the case of our selected data set. Furthermore, these extracted features are processed by the LSTM-based network in the second stage to exploit the temporal information. Finally, a single feature vector is obtained corresponding to each input video sequence, which is further fed to the dimension reduction and classification stages to reduce the feature

dimensions and then predict the class label, respectively. The comprehensive details of each stage are given in the succeeding sections.

Figure 2. Overall block diagram of our proposed spatiotemporal feature–based classification network composed of DenseNet and LSTM-based networks. KNN: k-nearest neighbor; LSTM: long short-term memory; PCA: principal component analysis.
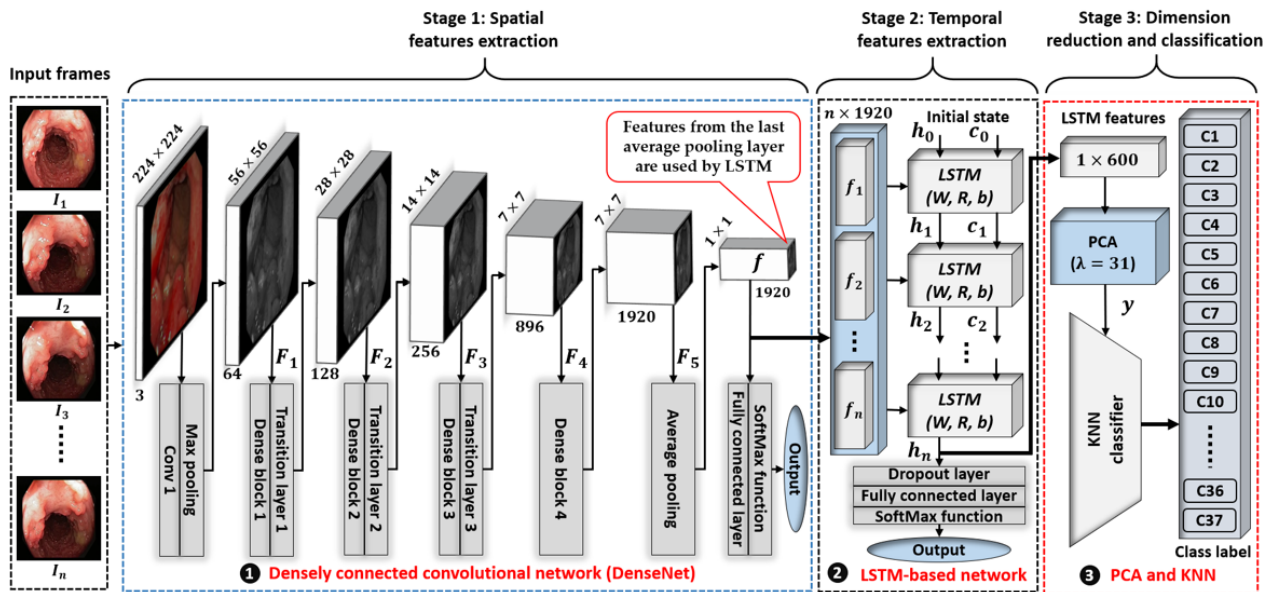
Spatial Feature Extraction Using DenseNet

The first stage of our proposed classification framework was composed of a well-known CNN architecture called DenseNet. Its main goal was to extract the spatial features from each frame of the given input video sequence independently. The main reason for selecting a DenseNet was that its classification performance is superior to other modern deep CNN models. Moreover, the dense connectivity in our selected model required fewer parameters than an equivalent traditional CNN. It also overcomes the vanishing gradient problem due to the presence of shortcut connectivity. During network training, the trainable parameters (ie, filter weights) receive an update proportional to the gradient value. In the case of the vanishing gradient problem, the gradient value becomes so small that it ultimately results in a small update and finally stops the training procedure. Such poor training significantly affects the overall performance of the network. However, the presence of shortcut connectivity in DenseNet overcomes this problem and results in better performance.

The complete layer-wise structure of the DenseNet is shown in Figure 2 and Table S2 in Multimedia Appendix 1. The entire network mainly comprises multiple dense blocks and transition layers, which could be considered the basic building blocks. There are a total of 4 dense blocks and 3 transition layers in it, and these make a significant contribution to exploiting discriminative features for better classification performance. The detailed structure of a single dense block followed by the transition layer is visualized in Figure S2 in Multimedia Appendix 1.

Each dense block is made up of multiple pairs of convolutional layers (conv 1×1 and conv 3×3) in sequential order. Furthermore, a feature concatenation layer is also added after each pair of convolutional layers (conv 1×1 and conv 3×3)

to concatenate the output feature maps of the current pair (conv 1×1 and conv 3×3) with the previous pair within the same dense block. In this way, each subsequent pair of convolutional layers directly accesses the output of all the previous pairs within the same block. A generalized expression to evaluate the output of the ℓth pair is given as follows:

$$x_\ell = H_\ell \left([x_0, x_1, x_2, \ldots, x_{\ell-1}] \right) \quad (1)$$

In equation 1, H_ℓ means the operation of the ℓth pair of convolutional layers (conv 1×1 and conv 3×3), which considers all the feature maps $(x_0, x_1, x_2, \ldots, x_{\ell-1})$ of the previous pairs within the same block. There is also another hyperparameter (labeled growth rate) in each dense block that regulates the increase in the depth of the output feature maps after passing through each pair of convolutional layers. Whereas the dimensions of the feature maps remain the same within each dense block, the number of filters between them changes. A generalized expression for the ℓth pair of convolutional layers (in each dense block) can be expressed as:

$$k_\ell = k_0 + k \times (\ell - 1) \quad (2)$$

In equation 2, k is the growth rate of the network; this hyperparameter was 32 in the original DenseNet model [27]. k_0 is the initial depth of the input feature map, x_0, that was passed to the dense block, and k_ℓ is the output depth of the ℓth pair of convolutional layers. Finally, the transition layer further processes the output feature map x_ℓ of the dense block and reduces its depth and dimension by passing it through convolutional and average pooling layers sized 1×1 and 2×2 pixels, respectively.

The given structural and parametric details (Table S2 in Multimedia Appendix 1) further illustrate the flow of spatial feature extraction through the different layers of the network. Initially, the first input layer (labeled "Image Input") in the

DenseNet is used to pass the given input image to the network for further processing. After the input layer, the first convolutional layer (labeled "Conv1") exploits the input frame by applying a total of 64 different filters sized $7 \times 7 \times 3$ pixels. The output feature map (generated by Conv1) is then processed by a max-pooling layer, which generates a down-sampled feature map, F_1, sized $56 \times 56 \times 64$ pixels. After the max-pooling layer, the first stack of the dense block and transition layer (labeled "Dense Block 1" and "Transition Layer 1") processes the feature map F_1 and generates a down-sampled feature map, F_2, sized $28 \times 28 \times 128$ pixels. The output feature map F_2 is further processed by the second stack of the dense block and transition layer (labeled "Dense Block 2" and "Transition Layer 2"), and an output feature map, F_3, sized $14 \times 14 \times 125$ pixels is obtained. Similarly, the third stack of the dense block and transition layer (labeled "Dense Block 3" and "Transition Layer 3") also processes the feature map F_3 and generates an output feature map, F_4, sized $7 \times 7 \times 896$ pixels. Output feature map F_4 is further processed by the last dense block (labeled "Dense Block 4") and produces an output feature map, F_5, sized $7 \times 7 \times 1920$ pixels. Finally, a spatial feature vector f sized $1 \times 1 \times 1920$ pixels is obtained after applying the last average pooling layer (labeled "Avg Pooling") with a filter size of 7×7 pixels over the last output feature map, F_5. The same procedure is repeated for all the other input frames, which ultimately generates a set of n feature vectors ($f_1, f_2, f_3 \ldots, f_n$) for all the successive endoscopic images ($I_1, I_2, I_3 \ldots, I_n$). Finally, all these feature vectors are further processed by the second-stage network for temporal feature extraction. There are also 3 other layers (labeled "fully connected," "softmax," and "classification output") after the last average pooling layer, as shown in Figure 2. These 3 layers only take part in the spatial training procedure of the DenseNet. Therefore, after completing the training phase, the final spatial features are selected from the last average pooling layer and are further processed in subsequent stages.

Temporal Feature Extraction Using an LSTM Network

In this second stage, an LSTM-based neural network (a version of a recurrent neural network) [28] is used to learn the temporal features from the spatial features (extracted in the first stage). In the case of temporal information-based challenges, the LSTM-based neural networks overcome the vanishing gradient problem, which causes poor training of a network due to a small gradient value. This vanishing gradient problem occurs through the repeated use of a recurrent weight matrix in a recurrent neural network. However, this problem is resolved in LSTM by replacing the recurrent matrix with the identity function. Therefore, in our proposed classification framework, a simplified structure of an LSTM-based network is adopted to further enhance classification performance by extracting the temporal features. A complete network structure and layer-wise description, including the parametric details, are given in Figure 2 (stage 2) and Table S2 in Multimedia Appendix 1, respectively. In Figure 2, the LSTM block presents a standard LSTM cell, which can be considered the main building block of our proposed LSTM-based network. In a standard LSTM cell, 3 different types of learnable parameters (recurrent weights R, input weights W, and bias b) are involved, which are trained

using the training data set. These learnable parameters (W, R, b) are responsible for learning the temporal features from the given training data set. Complete details about the internal structure of the LSTM cell are provided in Hochreiter and Schmidhuber [29]. For a better understanding, an unrolled version of the LSTM cell is shown in the second-stage network of Figure 2, which presents n executions of a single LSTM cell. Here, the parameter n (number of executions of a single LSTM cell) is a variable directly related to the length of a given input sequence, which provides the flexibility to classify input sequences with different numbers of frames ($I_1, I_2, I_3 \ldots, I_n$).

The extracted set of n spatial features ($f_1, f_2, f_3 \ldots, f_n$) in the previous stage is processed by this second stage network in sequential order, which can be visualized in Figure 2 (stage 2). A sequence input layer (Table S2 in Multimedia Appendix 1) is used to pass these spatial features to the LSTM layer, which is composed of multiple LSTM cells with different input parameters. In an actual scenario, a single LSTM cell is repeated n times to process the set of n spatial features ($f_1, f_2, f_3 \ldots, f_n$) in sequential order using the state information (hidden state h_{n-1} and cell state c_{n-1}) of all the previous input feature vectors (f_1 to f_{n-1}). The hidden state h_{n-1} holds the output of the LSTM cell for the input feature f_{n-1}, and the cell state c_{n-1} keeps the information learned from all the previous input feature vectors, f_1 to f_{n-1}. In the case of the first input feature vector, f_1, the LSTM cell considers the initial state of the network to be null values ($h_0 = []$, $c_0 = []$) when computing the first updated cell state c_1 and output h_1. For all the succeeding input feature vectors ($n \neq 1$), the LSTM cell considers the current state of the network (h_{n-1}, c_{n-1}) to compute the output h_n and the updated cell state c_n. This way, after processing all the input feature vectors ($f_1, f_2, f_3 \ldots, f_n$), the last hidden state h_n of the network is considered to be the final output feature vector for performing further dimension reduction and classification. There are also 4 other layers (labeled dropout, fully connected, softmax, and classification output) after the LSTM layer, as mentioned in Table S2 in Multimedia Appendix 1. These 4 layers only take part in the training procedure for this second stage network. Therefore, after training, the final features are selected from the LSTM layer for further processing in the next stage.

Dimension Reduction and Classification

Because the last hidden state h_n of the network (with a feature dimension of 1×600 pixels) includes the complete spatiotemporal information for all the input feature vectors ($f_1, f_2, f_3 \ldots, f_n$), it was therefore selected as the final output feature vector for classification. However, before applying the classification algorithm, a PCA was performed to further reduce the dimension of the final output feature vector h_n by using MATLAB R2019a (MathWorks Inc) [30]. This step was taken to reduce the feature comparison time for retrieval purposes and improve the overall classification accuracy of the proposed classification network. Therefore, an intermediate features-based data set (as shown in Figure S3 in Multimedia Appendix 1 after the LSTM-based network) was created for all the training and testing samples by extracting the output features from the last

hidden state h_n of the LSTM-based network. This newly obtained data set (in terms of feature vectors) was used to perform dimension reduction using PCA. The overall average performance corresponding to the different number of eigenvectors (λ=1,2,3, ...,600) was evaluated to select the optimal number of eigenvectors (λ). We obtained the best average performance for λ=31. A final set of feature vectors (with a feature dimension of 1×31 pixels) was created for all the training and testing data sets. Figure S3 in Multimedia Appendix 1 shows the conceptual representation of all the intermediate data sets created by our proposed classification framework at different stages. In Figure S3 in Multimedia Appendix 1, *k* presents the total number of data samples in the entire data set (including both training and testing).

Finally, the KNN [31] algorithm was applied using MATLAB R2019a [30] to classify this newly obtained features-based data set after the PCA. This simple classification algorithm was selected based on its classification performance relative to other classification algorithms, such as adaptive boosting (AdaBoostM2) [32] and a multiclass support vector machine (SVM) (Multi-SVM) [33]. It predicts the class label for the given testing sample by calculating the distance to the different neighbor samples and selecting the neighbor with the minimum distance. In our case, there were a total of 37 different categories related to the human GI tract, including both normal and abnormal cases. Therefore, the KNN algorithm finds the best class prediction for the given input testing data sample by identifying the nearest neighbor (based on Euclidean distance) of the 37 different neighbors. Finally, the predicted class label

is assigned to the given testing data sample, which was the ultimate objective of our proposed classification model.

Results

Data Set and Preprocessing

We evaluated the performance of our proposed classification network on 2 publicly available endoscopic databases: a dataset from Gastrolab [34] and the Kvasir dataset [35]. As the databases state [34,35], these are open databases and can be used for research purposes. These databases consist of various endoscopic videos [34] and some already extracted frames (including both normal and disease cases) [35] related to different anatomical districts of the human GI tract. From these 2 databases [34,35], we collected a total of 77 endoscopic videos with a total of 52,471 frames, as described in our previous work [10]. For each video, the anatomical districts, disease type, and other related details were provided in the video title. Based on the available information related to each video, we categorized the entire data set into 37 different classes, which encompassed the 5 main anatomical districts of the human GI tract, labeled the esophagus, the stomach, the small intestine, the large intestine, and the rectum. Figure 3 presents a pictorial representation of these different anatomical districts of the human GI tract and their corresponding groups of classes, with sample frames for each class (ie, C1 to C37). There are also other organs in the human GI tract, but the available data cover only the described 5 anatomical districts and include both normal and disease cases. Our proposed classification network shows the best performance for this data set, and it is also capable of classifying a large number of classes.

Figure 3. Example frames of our selected data set for each class and their correspondence with the different anatomical districts of the human gastrointestinal tract.
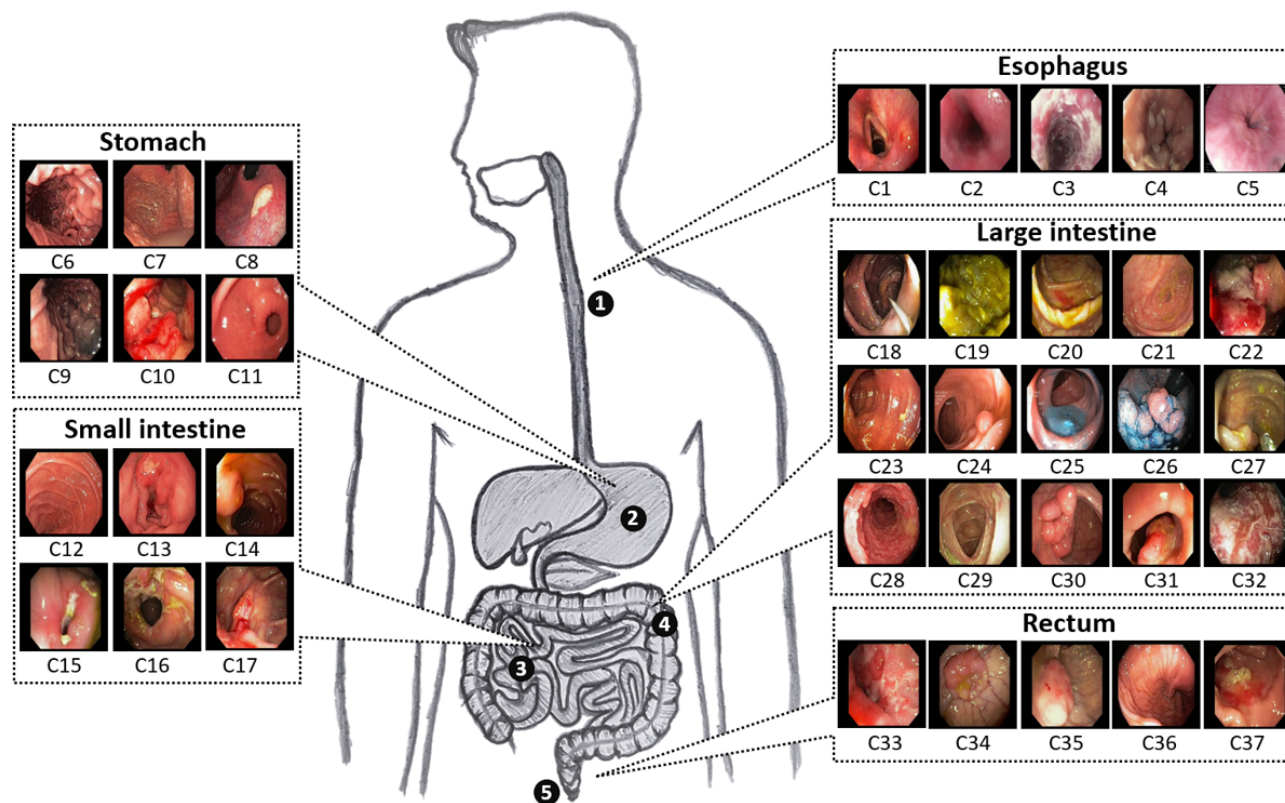
In addition, our selected data set shows high intraclass variance due to the varying textures and structures of the specific types of GI disease, such as tumors or cancer. Moreover, the dynamic structural changes during the endoscopy procedure and the different viewing conditions may also cause high intraclass variance. Figure S4 in Multimedia Appendix 1 shows a few example frames from our selected data set that illustrate this high intraclass variance. In this situation, it is difficult to capture a generalized feature-level abstraction that can represent all the possible samples of a class. However, the high intraclass variance in the data set may help analyze the performance of our proposed network in a more challenging scenario. The problem caused by the intraclass variance can be solved using a versatile and sufficient amount of training data related to each class. Such types of data may help extract a generalized high-level representation for each class by training a deep learning–based network.

Furthermore, Table S3 in Multimedia Appendix 1 provides additional details about the subcategories of each anatomical district and their corresponding classes and includes the actual class description and number of training and testing frames. Due to the different spatial resolutions, we resized all the extracted frames to a spatial dimension of 224×224 pixels (as the input layer size of the first-stage network) and then converted them to a BMP file format. We performed 2-fold cross-validation in all the experiments by dividing the entire data set, using 50% of the data for training and the remaining 50% for testing. The data splitting was performed by considering the first half of consecutive frames of a video as training data and the remaining half as testing data. Due to the limited size of the data set, it was not possible to use different videos in training and testing. In addition, most of the classes (about 21 classes out of 37 classes) consisted of single-patient data (ie, one video per class), as mentioned in Table S3 in Multimedia Appendix 1. Although we used the former part of a video for training and the latter for testing, the data were completely different due to the movement of the capturing device and the body organs, as shown in Figure S5 in Multimedia Appendix 1. In Figure S5 in Multimedia Appendix 1, a significant visual difference between the training and the testing data can be observed for some classes with single-patient data (ie, one video per class). Nevertheless, to highlight the superiority of our method, we also evaluated the results of some existing state-of-the-art methods [5,8,10,36-47] based on the same data set and experimental protocol. Additionally, online data augmentation [48] (with random rotation and translation in both directions) was applied (only in training the first-stage network) to resolve the class imbalance problem [49]. This class imbalance problem occurred due to the different amounts of training samples in each class; therefore, we applied online data augmentation only for the training data set.

Experimental Setup

The proposed classification and retrieval framework was developed using a deep learning toolbox from MATLAB R2019a [30]. This toolbox provides a comprehensive framework for designing and implementing various types of ANNs, including pretrained networks. All the experiments were conducted using a standard desktop computer with a 3.50 GHz Intel Core i7-3770K central processing unit [50] and 16 GB

RAM, an NVIDIA GeForce GTX 1070 graphics card [51], and the Windows 10 operating system. Both networks (stage 1 and stage 2) were trained using a well-known backpropagation algorithm called stochastic gradient descent [17]. Its main objective is to find the optimal parameters (ie, filter weights and biases) of a network iteratively by optimizing an objective function. Moreover, due to the limited data set, the initial parameters of the first-stage network were initialized with the filter weights of a pretrained CNN model (DenseNet201), which was already trained on the ImageNet data set [52]. This is a valid transfer learning approach that effectively converges the network training process, especially in the case of a small training data set. Similarly, the initial parameters of the second-stage network were assigned randomly by choosing Gaussian distribution with a mean of 0 and standard deviation of 0.001. Comprehensive details of all the training parameters are given in Table S4 in Multimedia Appendix 1.

After setting up the experimental protocol, the sequential training of the DenseNet (stage 1) and then the LSTM-based network (stage 2) was performed using the training data set. In the first stage, the DenseNet was trained using our selected data set (extracted endoscopic frames from the videos) to extract the spatial features. The progress of the training accuracy and loss according to the different numbers of epochs is visualized in Figure S6 in Multimedia Appendix 1 (for both folds of cross-validation). As seen in the figure, after increasing the number of epochs, the training accuracy approached 100% and the training loss decreased to 0, which shows that our network has been trained. In the second stage, the LSTM-based network was trained to further extract the temporal information using the spatial features–based data set (extracted from the DenseNet after completing the training). In this stage, the n successive feature vectors (extracted from the n successive frames in the first stage) were considered to be one training sample. Therefore, we converted each training sample sized 1×1920 pixels to $n \times 1920$ pixels by embedding the previous $n - 1$ feature vectors (extracted from the $n - 1$ consecutive frames). This newly obtained data set was used to train the second-stage network to further exploit the temporal information. We considered the optimal number of successive frames for training to be 14 (ie, n=14) after performing a number of experiments. The subsequent sections provide a detailed explanation for the experimental results for different values of n. Figure S7 in Multimedia Appendix 1 visualizes the training progress (in terms of training loss and accuracy) of the LSTM-based network according to different numbers of epochs. As the figure shows, the network convergence speed is faster and smoother than DenseNet. The main reason for this is the use of a spatial features–based data set rather than the original frames in this second-stage network for further temporal feature extraction. Finally, we assessed the performance of our proposed classification network using 4 quantitative evaluation metrics: accuracy, F1 score, mean average precision (mAP), and mean average recall (mAR) [53]. These are the most common measures for calculating the overall performance of a network from all perspectives.

Testing of the Proposed Method

The optimum number of successive frames (n) plays a vital role in exploiting temporal information, which ultimately results in better classification performance. A small value of n incorporates fewer temporal features, while a high value of n increases the effect of noise and the processing time. Therefore, it was necessary to find the optimal number of successive frames. For this purpose, we assessed the overall performance of our proposed classification network by considering different numbers of successive frames (ie, $n = 1,2,3, …,20$) in both the training and testing phase. The ultimate objective of these experiments was to find the value of n that showed the best performance. Figure S8 in Multimedia Appendix 1 shows the average performance of our proposed network for different values of n. We obtained the best average performance (as highlighted with the green square box in Figure S8 in Multimedia Appendix 1) for n=14. Therefore, we considered n=14 to be the optimal training parameter that would exploit temporal information while achieving significant performance gain, and we performed all the other experiments with this parameter setting.

After selecting the optimal value of n for the training of the LSTM-based network, the performance of our trained network (for the optimal value of n) was also evaluated by considering different numbers of successive frames (ie, $n = 1,2,3, …,150$). Figure S9 in Multimedia Appendix 1 shows the average classification performance for different values of n (ie, $n = 1,2,3, …,150$) used only in the testing stage. These results demonstrated that the overall performance of our proposed network was directly proportional to the number of successive frames (ie, $n = 1,2,3, …,150$) selected in the testing phase. The main reason for this is that a greater number of successive frames encompass more temporal information, which results in better classification performance. All the performance metrics showed a similar performance gain, which illustrated the significance of our proposed network compared with conventional deep CNN models. However, it was observed that the overall change in performance gain became smaller as the number of successive frames (ie, $n = 1,2,3, …,150$) increased. In this performance analysis setup, we obtained the highest performance for n=146 (as highlighted with a green square box in Figure S9 in Multimedia Appendix 1) instead of higher values of n. Therefore, we considered n=146 to be the testing parameter value for our selected data set. Both parameters that showed the best results for the various recent ANN methods (n=14 for training and n=146 for testing) were selected for the complete network.

To further enhance the prediction capability of the proposed network, we performed additional experiments by applying PCA followed by KNN after the LSTM-based network (as explained in the "Methods" section). In this performance analysis setup, we evaluated the PCA-based performance for the different numbers of eigenvectors ($\lambda = 1,2,3, …600$), as shown in Figure S10 in Multimedia Appendix 1. These results were computed to find the number of eigenvectors (λ) that showed the best performance. Of all the performance results, we found the maximum average performance for λ=31, as highlighted with a green square box in Figure S10 in Multimedia

XSL•FO

RenderX

Appendix 1 (left side), which presents a close-up view to further magnify the performance difference. Detailed comparative results (with and without PCA) are also given in Table 1 to show the effect of PCA on the LSTM-based network. In this analysis, the comparative results without PCA are based on a fully connected network (comprising fully connected, softmax, and classification output layers after the LSTM layer). As seen in Table 1, the PCA-based classification performance was higher than the fully connected network (without PCA). Furthermore, the dimension of the PCA-based feature vectors (1×31 pixels) is about 20 times lower than the original feature vectors (1×600 pixels), which also results in better retrieval performance. Consequently, our proposed classification network (including the PCA followed by KNN after the LSTM-based network) showed superior performance in all respects.

Table 1. Performance comparisons of our proposed network with and without applying the PCA.

Fold	Performance without PCA[a] (using fully connected network), %				Performance with PCA + KNN[b] (λ=31), %			
	Accuracy	F1	mAP[c]	mAR[d]	Accuracy	F1	mAP	mAR
Fold 1	97.31	97.83	97.97	97.70	98.41	98.80	98.82	98.79
Fold 2	94.18	95.11	96.91	93.38	93.97	95.18	97.54	92.94
Average, mean (SD)	95.75 (2.21)	96.47 (1.92)	97.44 (0.75)	95.54 (3.05)	96.19 (3.13)	96.99 (2.56)	98.18 (0.90)	95.86 (4.13)

[a]PCA: principal component analysis.

[b]KNN: k-nearest neighbor.

[c]mAP: mean average precision.

[d]mAR: mean average recall.

The detailed performance results of each class are shown in a confusion matrix in Figure 4. Each diagonal value in Figure 4 presents the individual class performance for accuracy. Moreover, the column (on the right side) and row (on the bottom) present the individual class performance for recall and precision, respectively. As seen in Figure 4, most of the classes show notably high classification performance (accuracy of at least 94%), but 3 classes show low performance (ie, C16, C31, and C33, which show average accuracies of 56%, 64%, and 55%, respectively). Such performance degradation is caused by the existence of high interclass similarity among the multiple classes. However, the overall performance of our proposed network was substantially high for a data set with high intraclass variance and high interclass similarity.

Figure 4. Detailed performance results of our proposed network shown in an average confusion matrix.
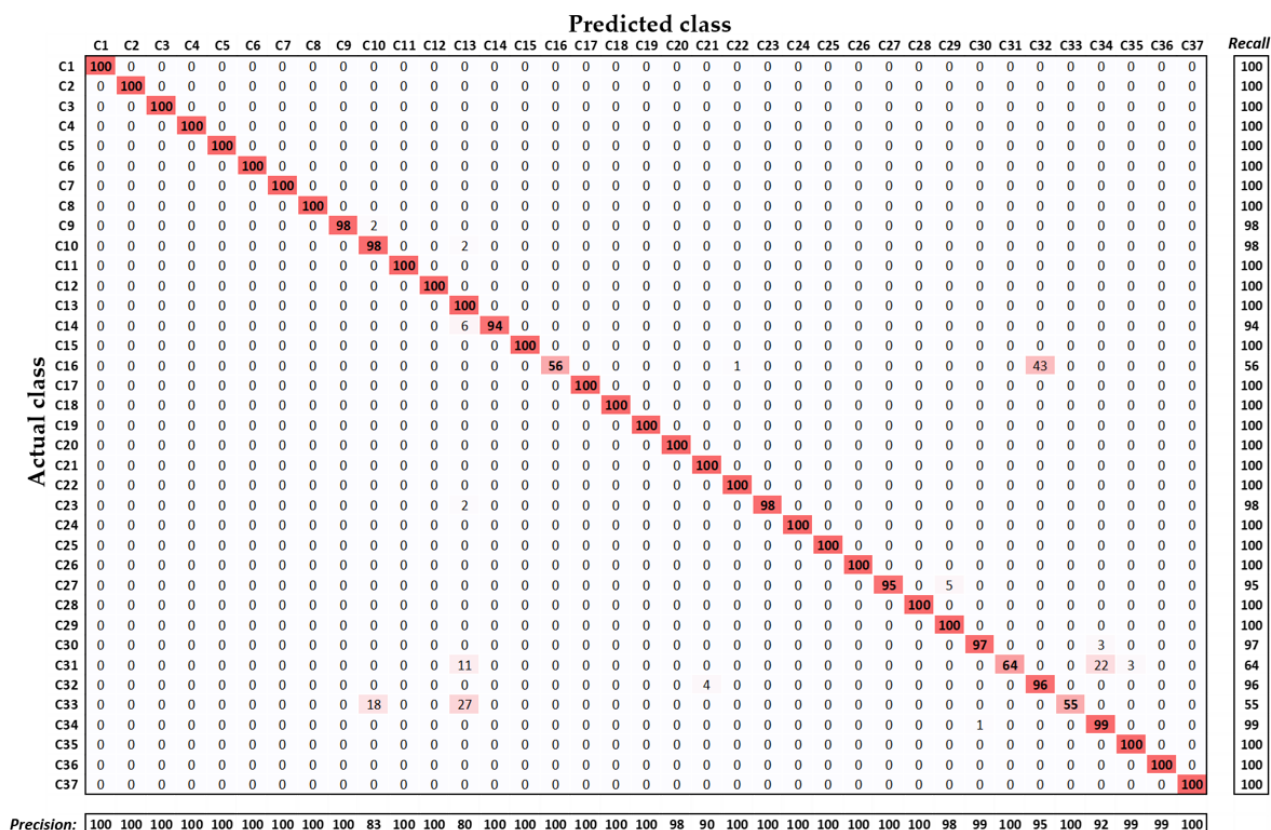
Comparisons With Previous Methods

We conducted a detailed performance comparison of our proposed method with other state-of-the-art deep learning methods for the automated detection of different types of GI diseases [5,8,10,36-47]. To make a fair comparison, the performance of these existing baseline methods was evaluated under the same experimental setup and data set that was used for our proposed method. We evaluated the performance of a total of 15 different baseline methods to make a detailed comparison with our proposed method. In this regard, our comparative analysis was more comprehensive than those of these existing studies [5,8,10,36-47]. Table 2 provides the comparative results of all the baseline methods in comparison with our proposed method. We obtained an average performance gain of 3.62%, 3.58%, 3.60%, and 3.58% for the accuracy, F1, mAP, and mAR, respectively, compared with the second-best network [10]. Moreover, our method outperformed the original DenseNet201 [27] (third-best network, which was used in Song et al's study [43]), with a performance gain of 4.07%, 4.57%, 5.27%, and 3.93% for the accuracy, F1, mAP, and mAR, respectively. These results signify the diagnostic ability of our proposed network for endoscopy image analysis for better treatment of various GI diseases. To envision the computational complexity of our proposed method and the baseline methods, brief parametric details of all the models are provided in Table S5 in Multimedia Appendix 1. Although the number of training parameters of our proposed network is higher than those in the second-best [10] and third-best [43] methods, its substantial performance gain distinguished it from the other models. The other baseline models [5,16,54-56] with a large number of training parameters showed lower performance compared with our proposed network.

Table 2. Comparative classification performance of our proposed network with the other state-of-the-art methods used in endoscopy.

Authors	Deep network	Accuracy, %	F1, %	mAP[a], %	mAR[b], %
Zhang et al (2017) [44]	SqueezeNet [57]	77.84	76.74	76.77	76.73
Hicks et al (2018) [45]	VGG19 [54]	85.15	85.29	85.88	84.72
Fan et al (2018) [36]	AlexNet [16]	80.08	80.49	80.70	80.28
Takiyama et al (2018) [8]	GoogLeNet [58]	84.59	85.14	85.29	84.99
Byrne et al (2019) [5]	InceptionV3 [55]	87.92	88.45	87.87	89.05
Jani et al (2019) [46]	MobileNetV2 [59]	88.53	88.51	88.34	88.69
Lee et al (2019) [39]	ResNet50 [56]	89.55	90.60	90.70	90.50
Vezakis et al (2019) [40]	ResNet18 [56]	89.95	90.35	90.72	89.99
Owais et al (2019) [10]	CNN[c] + LSTM[d] [28,56]	92.57	93.41	94.58	92.28
Cho et al (2019) [42]	InceptionResNet [60]	84.78	84.53	84.15	84.92
Dif et al (2020) [38]	ShuffleNet [61]	89.63	89.14	88.67	89.63
Song et al (2020) [43]	DenseNet201 [27]	92.12	92.42	92.91	91.93
Guimarães et al (2020) [37]	VGG16 [54]	85.72	85.80	86.24	85.37
Hussein et al (2020) [41]	ResNet101 [56]	90.24	91.14	91.52	90.78
Klang et al (2020) [47]	Xception [62]	86.05	84.88	84.19	85.58
Proposed method	DenseNet + LSTM + PCA[e] + KNN[f]	96.19	96.99	98.18	95.86

[a]mAP: mean average precision.

[b]mAR: mean average recall.

[c]CNN: convolutional neural network.

[d]LSTM: long short-term memory.

[e]PCA: principal component analysis.

[f]KNN: k-nearest neighbor.

Furthermore, we iteratively analyzed and compared the sensitivity performance of our proposed method and all the baseline methods. For this purpose, an experimental setup known as a Monte Carlo simulation [63] was carried out, in which the testing performance of each model was evaluated by randomly selecting 20% of testing samples as an intermediate testing data set. The entire experiment was repeated a total of 200 times for both folds of the cross-validation (100 iterations for each fold), and we obtained 200 different performance results for each performance metric. We then calculated the average and standard deviation (for each metric), which presented the overall sensitivity performance of each method. Figure S11 in Multimedia Appendix 1 shows the comparative sensitivity results of our proposed method and the baseline methods. Based on these results, it can be concluded that the overall sensitivity performance of our proposed classification network was substantially higher than those of the existing networks.

We further performed a 2-tailed t test [64] and Cohen d [65] analysis to determine the significance of the performance gain

of our proposed method compared with the second-best [10] and third-best [43] baseline methods. The sensitivity results (obtained in the previous section) were used to evaluate the performance of the *t* test and Cohen *d* analysis quantitatively. Generally, a *t* test analysis is carried out to magnify the performance difference of two models or algorithms in a quantitative way using a null hypothesis (*H*), which assumes that two systems are similar (ie, *H*=0). A rejection score (*P* value) between the two systems is calculated, which ultimately gives a confidence score for the rejection of this null hypothesis.

In the Cohen *d* [65] analysis, the performance difference between two systems is determined by measuring effect size [66], which is normally categorized as small (approximately 0.2-0.3), medium (approximately 0.5), and large (≥0.8). A large effect size shows a significant performance difference between the systems. For our proposed method, we separately evaluated its rejection scores (*P* values) and effect sizes with the second-best and then the third-best baseline methods. The complete performance analysis results (for both the *t* test and Cohen *d*) are given in Table 3.

Table 3. The *t* test and Cohen *d* performance analysis results (*P* values and effect sizes).

Methods	Proposed vs second-best method				Proposed vs third-best method			
	Accuracy	F1	mAP[a]	mAR[b]	Accuracy	F1	mAP	mAR
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Cohen *d*	2.51	1.78	1.89	1.55	3.4	2.32	2.88	1.89

[a]mAP: mean average precision.

[b]mAR: mean average recall.

As seen in Table 3, the *P* values (*t* test analysis) are less than .001 for all performance metrics, which indicates that the null hypothesis is rejected (ie, *H*≠0), with a 99% confidence interval. Similarly, the effect sizes (Cohen *d* analysis) are higher than 0.8 for all the performance metrics. These 2 performance analysis results indicate that our proposed method shows a significant performance difference compared with both baseline methods. Moreover, Figure 5 presents the comparative performance (in mean, standard deviation, *P* value, and effect size) of our proposed method in comparison with the second-best and third-best models. The higher mean performance of our method (according to all the performance metrics) shows its superiority over the second-best and third-best baseline model.

Figure 5. The *t* test and Cohen *d* performance comparison of our method with the second-best and third-best models with the average accuracy, F1 score, mAP, and mAR. CNN: convolutional neural network; LSTM: long short-term memory; mAP: mean average precision; mAR: mean average recall.
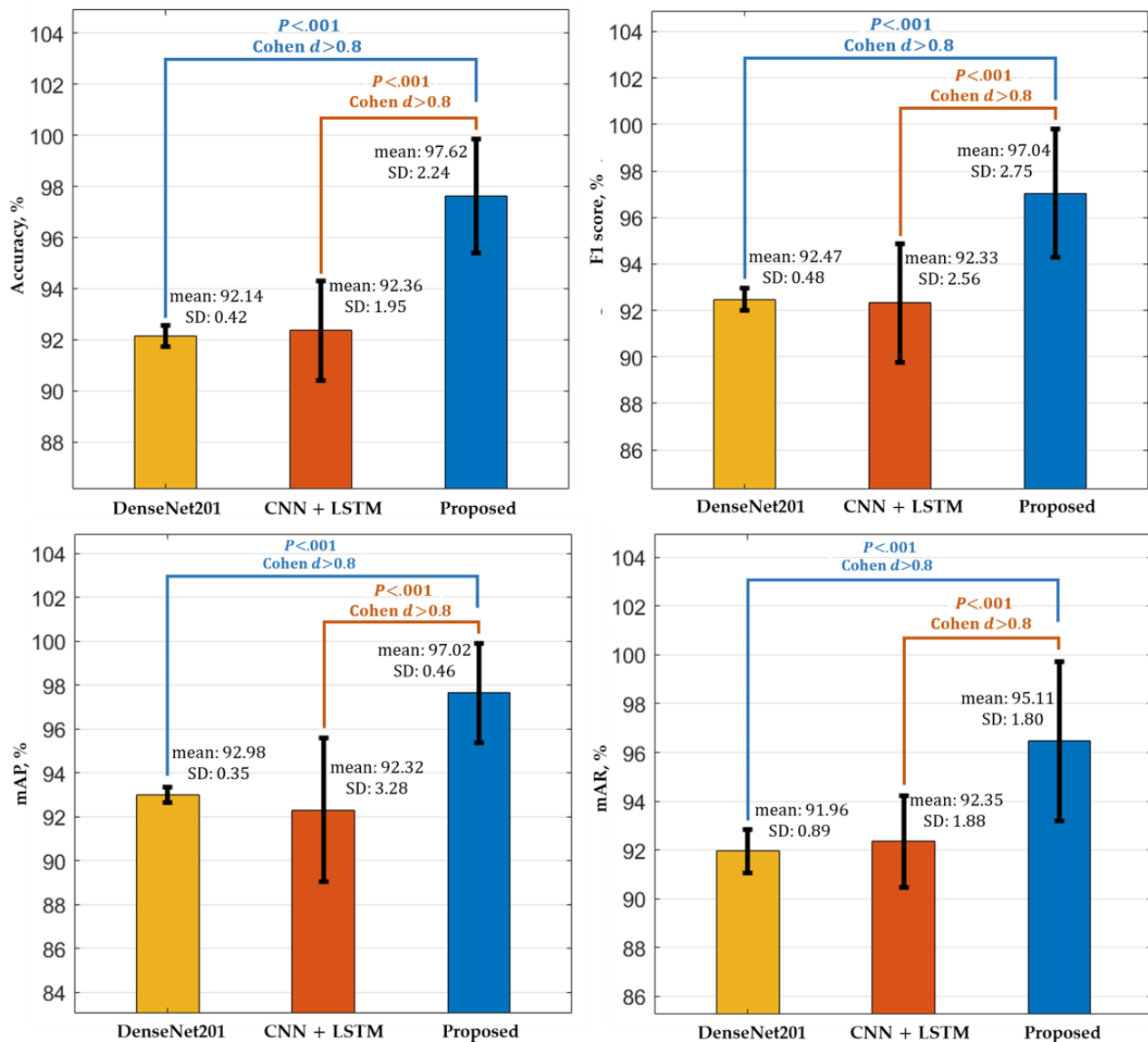
Finally, we made a detailed comparison of our proposed network with various handcrafted feature–based methods. We evaluated the performance of our selected data set using 3 conventional handcrafted feature extraction methods—local binary pattern (LBP) [67], histogram of oriented gradients (HOG) [68], and multilevel local binary pattern (MLBP) [69]—with 4 different classifiers—AdaBoostM2 [32], Multi-SVM [33], random forest (RF) [70], and KNN [31]. In total, we obtained the classification performance of 12 different methods, as shown in Table 4. Among all these methods, the HOG feature extractor followed by the RF classifier showed the highest performance, with an average accuracy of 61.41%, F1 score of 63.19%, mAP of 68.66%, and mAR of 58.55%. This means the HOG features extractor method exploits more distinctive low-level features

(ie, corners, blobs, or edges) than LBP or MLBP. In addition, the tree structure of the RF classifier further improved the classification decisions, which ultimately resulted in better classification performance. However, our proposed method performed substantially higher than the best handcrafted feature–based method (ie, HOG features extractor with RF classifier). In conclusion, our proposed network outperformed the various handcrafted and the deep features–based methods. We performed additional comparisons of KNN with more sophisticated classifiers, such as AdaBoostM2, Multi-SVM, and RF. In the case of our proposed network, KNN showed the best performance compared with AdaBoostM2, Multi-SVM, and RF, as shown in Table 4.

Table 4. Classification performance comparison of our proposed method with the other handcrafted feature–based methods.

Feature descriptor and classifier	Accuracy, %	F1, %	mAP[a], %	mAR[b], %
Local binary pattern [67]				
AdaBoostM2	35.74	27.70	35.74	22.61
Multi-SVM[c]	43.84	42.35	42.99	41.72
RF[d]	57.10	53.85	54.79	52.95
KNN[e]	50.46	47.36	46.86	47.87
Histogram of oriented gradients [68]				
AdaBoostM2	39.35	32.86	39.35	28.22
Multi-SVM	49.84	53.80	67.39	44.88
RF	61.41	63.19	68.66	58.55
KNN	53.20	54.68	58.41	51.45
Multilevel local binary pattern [69]				
AdaBoostM2	44.02	37.45	44.02	32.59
Multi-SVM	55.47	53.10	54.75	51.55
RF	61.40	57.57	59.08	56.13
KNN	55.40	52.20	52.06	52.33
Proposed feature descriptor (DenseNet + LSTM[f] +PCA[g])				
AdaBoostM2	93.39	93.66	94.35	92.98
Multi-SVM	95.50	96.43	97.98	94.96
RF	81.16	82.96	84.48	81.55
KNN	96.19	96.99	98.18	95.86

[a]mAP: mean average precision.

[b]mAR: mean average recall.

[c]SVM: support vector machine.

[d]RF: random forest.

[e]KNN: k-nearest neighbor.

[f]LSTM: long short-term memory.

[g]PCA: principal component analysis.

Discussion

Principal Findings

In this research, we used the strength of recent ANNs in endoscopy and proposed a high-performance CAD framework to diagnose multiple GI diseases simultaneously in a given endoscopic video. First, we implemented an efficient video classification network to classify endoscopic videos into one of 37 different categories (including both normal and diseased classes). Our proposed network contemplates both spatial and temporal features, which ultimately resulted in better performance in comparison with other modern classification networks. The spatial features are extracted in the first step using a densely connected convolutional network, and then an LSTM-based network further processes the spatial features to extract the temporal information. Therefore, optimal spatial features must be extracted to achieve the best performance by the LSTM-based network in the second stage. For this purpose,

we considered a DenseNet model, which shows superior classification performance compared with various CNN models. The performance difference between DenseNet and other CNN models can be observed in Table 2. These results show that our selected DenseNet model had a performance gain of 1.88% in average accuracy compared with the second-best CNN model (ResNet101). It exploits the optimal spatial features by processing the input frames through different dense blocks and transition layers, highlighting the class-specific discriminative regions [71] as the optimal feature maps. The presence of dense blocks and transitions layers in DenseNet makes it different from the other CNN models and results in superior performance. Figure 6 presents the progress of class-specific discriminative regions (activation maps) through the different parts of the first-stage network. For better interpretation, we calculated an average activation map for each layer and represented it with a pseudocolor scheme (red as the maximum value and blue as the minimum value) by overlaying it on the input frame. As shown in Figure 6, the class-specific discriminative regions (activation

maps F1, F2, …, F5) become more prominent after passing through the different layers of the network of Figure 2. Ultimately, we can obtain class-specific regions (activation map F5) that include the particular visual pattern for each class. These final class-specific activation maps are processed by the LSTM-based network for temporal feature extraction after passing through the global average pooling layer of the first-stage network. The key difference between DenseNet [27] and the LSTM-based network [10,28] is the extraction of 2 different types of features from the given input video sequence. The LSTM-based network exploits the time-dependent features of the successive frames of a video, while DenseNet only extracts the features within a single frame. Because of these 2 different types of features, the networks are different from each other. However, their combined connectivity generates an optimal representation of the given input video of n consecutive frames (ie, $I_1, I_2, I_3 …, I_n$) in terms of spatiotemporal features, which ultimately results in better classification performance. The significant performance gain by our proposed network boosts its usability in the diagnosis of various GI diseases by automatically detecting and classifying various types of GI diseases, such as gastric ulcers, cancer, or polyps. In the second step, these classification results were further used to retrieve relevant cases (endoscopic frames) from the database that are closely related to the current medical condition of a patient. Our proposed method is based on class prediction–based retrieval, in which feature matching is performed only within the predicted class to find the best matches. However, without class prediction–based retrieval, feature matching is done with the entire database, which is time-consuming.

Figure 6. Obtained class-specific discriminative regions from different parts of the first-stage network (DenseNet) for given input frames.
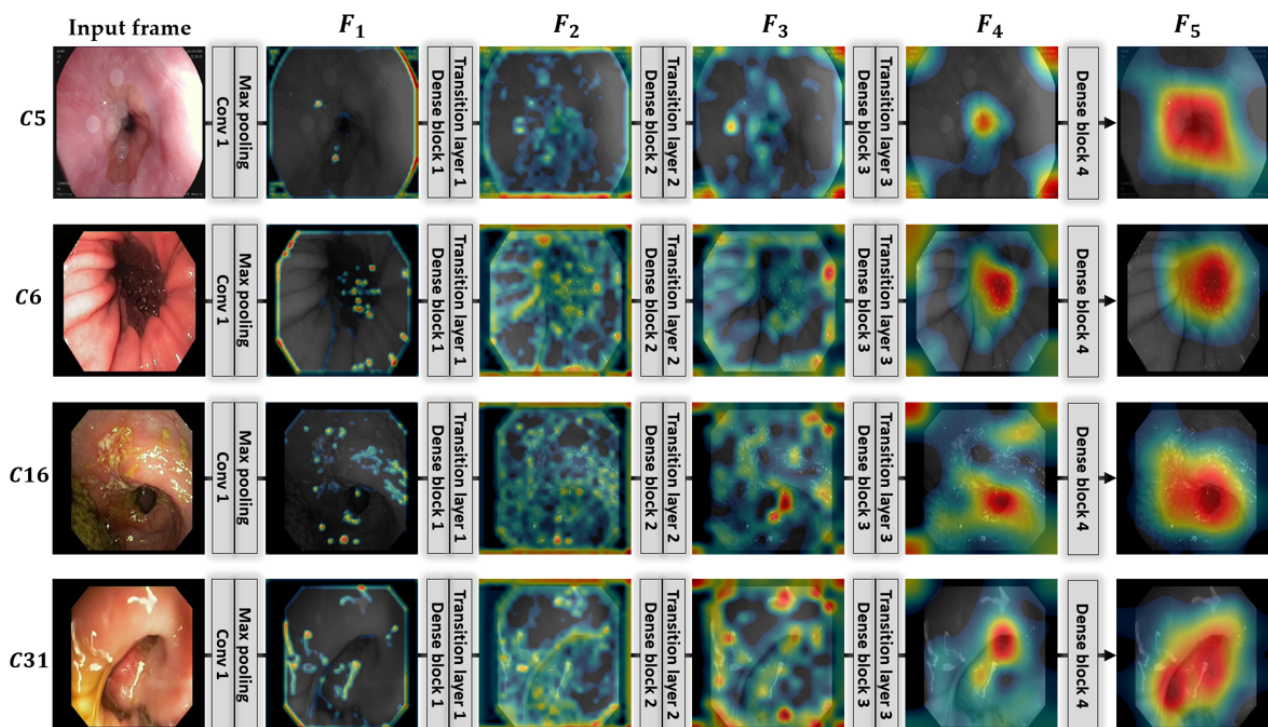
We computed the retrieval performance for both methods (with and without class prediction) using the proposed and second-best baseline model [10], as presented in Table 5. These comparative results showed a substantial performance difference between our proposed method and the second-best method for both cases (with and without class prediction). The retrieval performance of our proposed class prediction–based retrieval method was also better than the performance of the method without class prediction–based retrieval. In addition to the performance difference, the main advantage of our class prediction–based retrieval method is the optimal features comparison time required to retrieve the relevant cases from the database. In the method without class prediction, this feature comparison time was significantly higher because of the large number of feature comparisons, as the number of feature comparisons is directly related to the volume of the whole data set (ie, the total number of available data samples in the data set). Moreover, this feature comparison time also increases with the increasing number of data samples in each class. On the other hand, in our proposed class prediction–based retrieval method, there is no relation between the feature comparison time and the volume of the data set. Hence, our proposed retrieval method is more efficient in terms of retrieval computational cost.

Table 5. Performance comparisons of our proposed and the second-best baseline method [10] using both retrieval methods.

Method	With class prediction				Without class prediction			
	Accuracy, %	F1, %	mAP[a], %	mAR[b], %	Accuracy, %	F1, %	mAP, %	mAR, %
Owais et al [10]	92.57	93.41	94.58	92.28	93.18	94.02	94.68	93.38
Proposed	96.19	96.99	98.18	95.86	96.13	96.94	98.04	95.89

[a]mAP: mean average precision.

[b]mAR: mean average recall.

In addition, Figures 7 and 8 show the retrieved frames for the given input query from classes C16 and C31, respectively. All these results were computed separately for both retrieval methods (our proposed class prediction–based method and other methods without class prediction) to show the performance difference visually. All the results are presented in ranked order by retrieving the 24 best matches for each input query. As seen in the figures, our proposed method retrieves all the best matches as true positives for the given input query, whereas the methods without class prediction show many false-positive frames (highlighted with a red bounding box) among the retrieved frames. Hence, the class prediction–based retrieval method outperforms the other method in retrieving multiple best matches, and the retrieval methods without class prediction show performance degradation in retrieving multiple best matches.

Figure 7. Obtained retrieval results in ranked order (1st to 24th best matches) for the C16 input query frame using both a class prediction–based method and a method without class prediction.

Figure 8. Obtained retrieval results in ranked order (1st to 24th best matches) for the C31 input query frame using both a class prediction–based method and a method without class prediction.
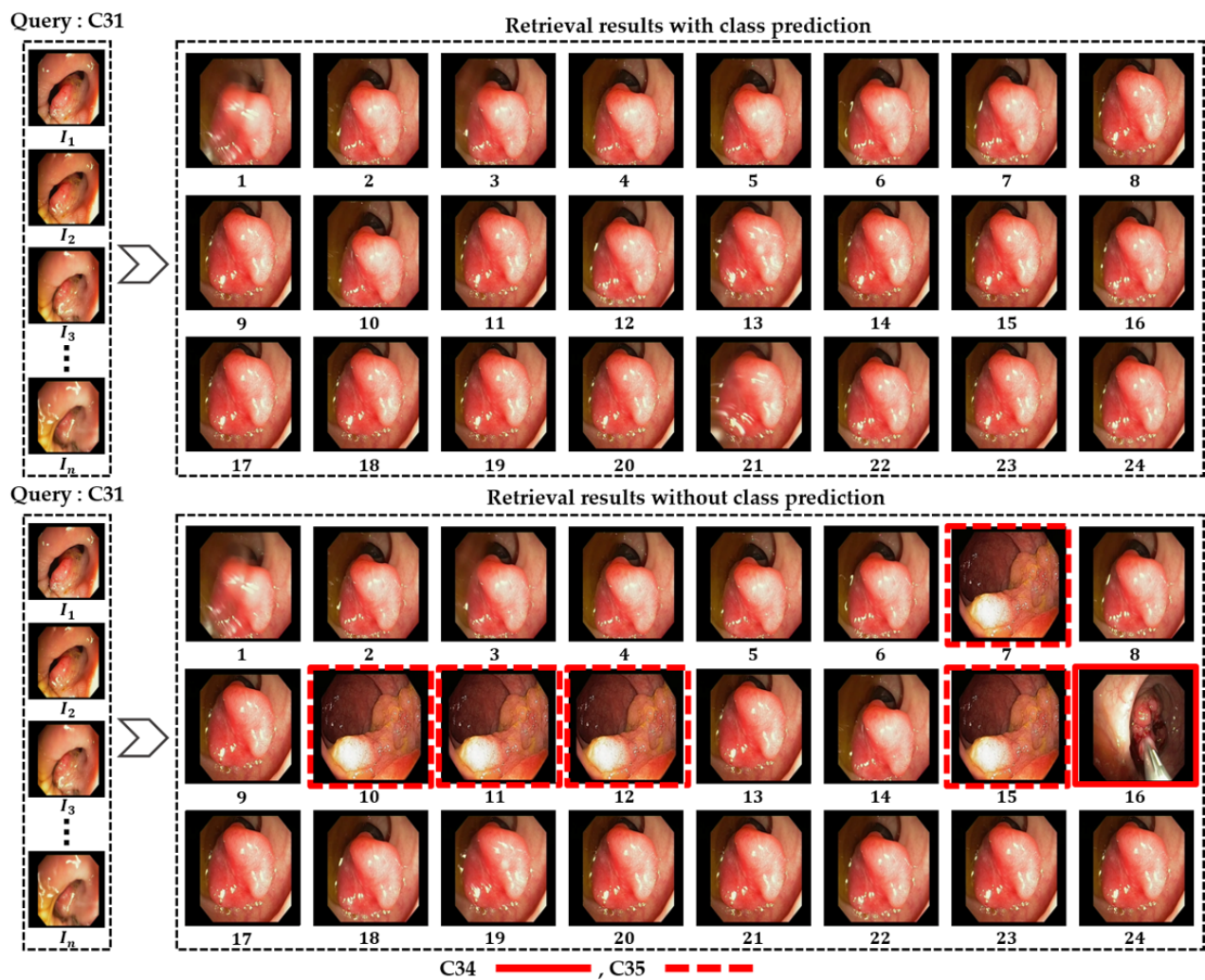
Limitations and Future Work

Due to a limited data set, data splitting was performed by considering the first half of the consecutive frames of a video as the training data and the remaining half as the testing data. Thus, the testing data set was obtained from the same sources as the training data set, which may raise issues of the generalizability of our framework. However, we accomplished our goal to incorporate a large number of GI diseases in a single deep learning–based CAD framework and provided an initial pretrained network in the field of GI diagnostics. To highlight the superiority of our proposed solution, we used a similar data splitting and experimental protocol to evaluate the results of various existing methods [5,8,10,36-47]. Finally, we provided a novel baseline solution in the emergent clinical setting as a supporting tool that can be further evolved in future studies.

According to our experimental analysis, a database of sufficient size with all of the common diseases can further enhance the generalizability of our proposed framework. In future work, we will further explore a large number of cases with this data set and try to further enhance the overall performance of the system by using different videos in training and testing.

Conclusions

This study presented a comprehensive CAD tool, a deep learning–based classification-driven retrieval framework, for identifying various types of GI diseases. The complete framework comprises a deep learning–based classification network followed by a retrieval method. The classification network predicts the disease type for the current medical condition, and the retrieval part then shows the relevant cases from the previous database. As a result, past cases help the medical expert subjectively validate the current prediction by the CAD method, which ultimately results in better diagnosis and treatment. In the case of a wrong prediction by the computer, the medical expert can check other relevant cases (ie, second-, third-, or fourth-best matches), which may be more relevant than the first-best match. Our results (also provided in Multimedia Appendix 2) show the superiority of our proposed method over various other state-of-the-art methods.

Acknowledgments

Authors' Contributions

MO and KRP designed the overall framework and wrote and revised the complete paper. MA, TM, and JKK helped design the comparative analysis and experiments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Other supplementary material.
[DOCX File , 4455 KB - jmir_v22i11e18563_app1.docx]

Multimedia Appendix 2
Experimental results.
[XLSX File (Microsoft Excel File), 213 KB - jmir_v22i11e18563_app2.xlsx]

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7-30 [FREE Full text] [doi: 10.3322/caac.21332] [Medline: 26742998]
2. Seguí S, Drozdzal M, Pascual G, Radeva P, Malagelada C, Azpiroz F, et al. Generic feature learning for wireless capsule endoscopy analysis. Comput Biol Med 2016 Dec 01;79:163-172. [doi: 10.1016/j.compbiomed.2016.10.011] [Medline: 27810622]
3. Zhou T, Han G, Li B, Lin Z, Ciaccio E, Green P, et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method. Comput Biol Med 2017 Jun 01;85:1-6. [doi: 10.1016/j.compbiomed.2017.03.031] [Medline: 28412572]
4. Zhang R, Zheng Y, Mak T, Yu R, Wong S, Lau J, et al. Automatic Detection and Classification of Colorectal Polyps by Transferring Low-Level CNN Features From Nonmedical Domain. IEEE J Biomed Health Inform 2017 Jan;21(1):41-47 [FREE Full text] [doi: 10.1109/jbhi.2016.2635662]
5. Byrne MF, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019 Jan;68(1):94-100 [FREE Full text] [doi: 10.1136/gutjnl-2017-314547] [Medline: 29066576]
6. Shichijo S, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, et al. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. EBioMedicine 2017 Nov;25:106-111 [FREE Full text] [doi: 10.1016/j.ebiom.2017.10.014] [Medline: 29056541]
7. Yu L, Chen H, Dou Q, Qin J, Heng P. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. IEEE J Biomed Health Inform 2017 Jan;21(1):65-75 [FREE Full text] [doi: 10.1109/jbhi.2016.2637004]
8. Takiyama H, Ozawa T, Ishihara S, Fujishiro M, Shichijo S, Nomura S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. Sci Rep 2018 May 14;8(1):7497 [FREE Full text] [doi: 10.1038/s41598-018-25842-6] [Medline: 29760397]
9. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018 Jul;21(4):653-660. [doi: 10.1007/s10120-018-0793-2] [Medline: 29335825]
10. Owais M, Arsalan M, Choi J, Mahmood T, Park K. Artificial Intelligence-Based Classification of Multiple Gastrointestinal Diseases Using Endoscopy Videos for Clinical Diagnosis. J Clin Med 2019 Jul 07;8(7):1-33 [FREE Full text] [doi: 10.3390/jcm8070986] [Medline: 31284687]
11. Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, et al. An Interpretable and Expandable Deep Learning Diagnostic System for Multiple Ocular Diseases: Qualitative Study. J Med Internet Res 2018 Nov 14;20(11):e11144 [FREE Full text] [doi: 10.2196/11144] [Medline: 30429111]

12.   Arsalan M, Owais M, Mahmood T, Cho S, Park K. Aiding the Diagnosis of Diabetic and Hypertensive Retinopathy Using Artificial Intelligence-Based Semantic Segmentation. J Clin Med 2019 Sep 11;8(9):1-28 [FREE Full text] [doi: 10.3390/jcm8091446] [Medline: 31514466]

13.   Ko H, Chung H, Kang W, Kim K, Shin Y, Kang S, et al. COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation. J Med Internet Res 2020 Jun 29;22(6):e19569 [FREE Full text] [doi: 10.2196/19569] [Medline: 32568730]

14.   Dhombres F, Maurice P, Guilbaud L, Franchinard L, Dias B, Charlet J, et al. A Novel Intelligent Scan Assistant System for Early Pregnancy Diagnosis by Ultrasound: Clinical Decision Support System Evaluation Study. J Med Internet Res 2019 Jul 03;21(7):e14286 [FREE Full text] [doi: 10.2196/14286] [Medline: 31271152]

15.   Faruqui S, Du Y, Meka R, Alaeddini A, Li C, Shirinkam S, et al. Development of a Deep Learning Model for Dynamic Forecasting of Blood Glucose Level for Type 2 Diabetes Mellitus: Secondary Analysis of a Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e14452 [FREE Full text] [doi: 10.2196/14452] [Medline: 31682586]

16.   Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. 2012 Presented at: 25th International Conference on Neural Information Processing Systems; Dec 3-6, 2012; Lake Tahoe, NV URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf [doi: 10.1145/3065386]

17.   Ruder S. An overview of gradient descent optimization algorithms. arXiv. 2017. URL: https://arxiv.org/abs/1609.04747 [accessed 2020-11-18]

18.   Barbosa D, Ramos J, Lima C. Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform. 2008 Presented at: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 20-25, 2008; Vancouver, BC, Canada URL: https://doi.org/10.1109/IEMBS.2008.4649837 [doi: 10.1109/iembs.2008.4649837]

19.   Karargyris A, Bourbakis N. Detection of Small Bowel Polyps and Ulcers in Wireless Capsule Endoscopy Videos. IEEE Trans Biomed Eng 2011 Oct;58(10):2777-2786 [FREE Full text] [doi: 10.1109/tbme.2011.2155064]

20.   Li B, Meng MQ. Automatic polyp detection for wireless capsule endoscopy images. Expert Syst Applications 2012 Sep;39(12):10952-10958. [doi: 10.1016/j.eswa.2012.03.029]

21.   Li B, Meng MQ. Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM-Based Feature Selection. IEEE Trans Inform Technol Biomed 2012 May;16(3):323-329 [FREE Full text] [doi: 10.1109/titb.2012.2185807]

22.   Manivannan S, Wang R, Trucco E, Hood A. Automatic normal-abnormal video frame classification for colonoscopy. 2013 Presented at: IEEE 10th International Symposium on Biomedical Imaging; April 7-11, 2013; San Francisco, CA URL: https://doi.org/10.1109/ISBI.2013.6556557 [doi: 10.1109/isbi.2013.6556557]

23.   Manivannan S, Wang R, Trucco E. Extended Gaussian-filtered local binary patterns for colonoscopy image classification. 2013 Presented at: IEEE International Conference on Computer Vision Workshops; Dec 2-8, 2013; Sydney, Australia URL: https://doi.org/10.1109/ICCVW.2013.31 [doi: 10.1109/iccvw.2013.31]

24.   Yuan Y, Li B, Meng MQ. Improved Bag of Feature for Automatic Polyp Detection in Wireless Capsule Endoscopy Images. IEEE Trans Automat Sci Eng 2016 Apr;13(2):529-535. [doi: 10.1109/tase.2015.2395429]

25.   Dongguk University. URL: http://dm.dgu.edu/link.html [accessed 2020-11-18]

26.   Ilin A, Raiko T. Practical approaches to principal component analysis in the presence of missing values. J Mach Learn Res 2010 Jul;11(1):1957-2000 [FREE Full text]

27.   Huang G, Liu Z, van der Maaten L, Weinberger K. Densely connected convolutional networks. 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; July 21-26, 2017; Honolulu, HI URL: http://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html [doi: 10.1109/cvpr.2017.243]

28.   Schuster M, Paliwal K. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997 Nov;45(11):2673-2681 [FREE Full text] [doi: 10.1109/78.650093]

29.   Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]

30.   Deep learning toolbox. Mathworks. URL: https://in.mathworks.com/products/deeplearning.html [accessed 2020-11-18]

31.   Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967 Jan;13(1):21-27 [FREE Full text] [doi: 10.1109/tit.1967.1053964]

32.   Freund Y, Schapire R. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J Comp Syst Sci 1997 Aug;55(1):119-139 [FREE Full text] [doi: 10.1006/jcss.1997.1504]

33.   Hsu C, Lin C. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 2002;13(2):415-425. [doi: 10.1109/72.991427] [Medline: 18244442]

34.   Gastrolab - the gastrointestinal site. Gastrolab. URL: http://www.gastrolab.net/ni.htm [accessed 2020-11-18]

35.   Pogorelov K, Randel K, Griwodz C, Eskeland S, de Lange T, Johansen D. KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection. 2017 Presented at: 8th ACM Multimedia Systems Conference; June 20-23, 2017; Taipei, Taiwan URL: https://doi.org/10.1145/3083187.3083212 [doi: 10.1145/3083187.3083212]

36.   Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys Med Biol 2018 Aug 10;63(16):165001. [doi: 10.1088/1361-6560/aad51c] [Medline: 30033931]

XSL•FO

RenderX

37. Guimarães P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. Gut 2020 Jan;69(1):4-6. [doi: 10.1136/gutjnl-2019-319347] [Medline: 31375599]

38. Dif N, Elberrichi Z. A new deep learning model selection method for colorectal cancer classification. Int J Swarm Intell Res 2020;11(3):88. [doi: 10.4018/ijsir.2020070105]

39. Lee J, Kim Y, Kim Y, Park S, Choi Y, Kim Y, et al. Spotting malignancies from gastric endoscopic images using deep learning. Surg Endosc 2019 Nov;33(11):3790-3797. [doi: 10.1007/s00464-019-06677-2] [Medline: 30719560]

40. Vezakis I, Toumpaniaris P, Polydorou A, Koutsouris D. A novel real-time automatic angioectasia detection method in wireless capsule endoscopy video feed. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 23-27 URL: https://doi.org/10.1109/EMBC.2019.8857445 [doi: 10.1109/embc.2019.8857445]

41. Hussein M, Gonzalez-Bueno Puyal J, Brandao P, Toth D, Sehgal V, Everson MA, et al. Sa2030 Deep Neural Network For The Detection Of Early Neoplasia In Barrett's Oesophagus. Gastrointest Endosc 2020 Jun;91(6):AB250 [FREE Full text] [doi: 10.1016/j.gie.2020.03.1826]

42. Cho B, Bang C, Park S, Yang Y, Seo S, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019 Dec;51(12):1121-1129. [doi: 10.1055/a-0981-6133] [Medline: 31443108]

43. Song E, Park B, Ha C, Hwang S, Park S, Yang DH, et al. Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model. Sci Rep 2020 Jan 08;10(1):30 [FREE Full text] [doi: 10.1038/s41598-019-56697-0] [Medline: 31913337]

44. Zhang X, Hu W, Chen F, Liu J, Yang Y, Wang L, et al. Gastric precancerous diseases classification using CNN with a concise model. PLoS One 2017;12(9):e0185508 [FREE Full text] [doi: 10.1371/journal.pone.0185508] [Medline: 28950010]

45. Hicks S, Eskeland S, Lux M, Lange T, Randel K, Jeppsson M. Mimir: an automatic reporting and reasoning system for deep learning based analysis in the medical domain. 2018 Presented at: 9th ACM Multimedia Systems Conference; June 12-15, 2018; Amsterdam, Netherlands p. 12-15 URL: https://doi.org/10.1145/3204949.3208129 [doi: 10.1145/3204949.3208129]

46. Jani K, Srivastava R, Srivastava S, Anand A. Computer aided medical image analysis for capsule endoscopy using conventional machine learning and deep learning. 2019 Presented at: 7th International Conference on Smart Computing & Communications (ICSCC); June 28-30, 2019; Sarawak, Malaysia URL: https://doi.org/10.1109/ICSCC.2019.8843669 [doi: 10.1109/icscc.2019.8843669]

47. Klang E, Barash Y, Margalit R, Soffer S, Shimon O, Albshesh A, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. Gastrointest Endosc 2020 Mar;91(3):606-613.e2. [doi: 10.1016/j.gie.2019.11.012] [Medline: 31743689]

48. Peng X, Tang Z, Yang F, Feris R, Metaxas D. Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation. 2018 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT URL: https://doi.org/10.1109/CVPR.2018.00237 [doi: 10.1109/cvpr.2018.00237]

49. Wong S, Gatt A, Stamatescu V, McDonnell M. Understanding data augmentation for classification: When to warp? 2016 Presented at: IEEE International Conference on Digital Image Computing: Techniques and Applications; Nov 30-Dec 2, 2016; Gold Coast, Australia URL: https://doi.org/10.1109/DICTA.2016.7797091 [doi: 10.1109/dicta.2016.7797091]

50. Intel Core i7-3770K Processor. Intel. URL: https://ark.intel.com/content/www/us/en/ark/products/65523/intel-core-i7-3770k-processor-8m-cache-up-to-3-90-ghz.html [accessed 2020-11-18]

51. GeForce GTX 1070. GeForce. URL: https://www.geforce.com/hardware/desktop-gpus/geforce-gtx-1070/specifications [accessed 2020-11-18]

52. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 20-25, 2009; Miami, FL URL: https://doi.org/10.1109/CVPR.2009.5206848 [doi: 10.1109/cvprw.2009.5206848]

53. Hossin M, Sulaiman MN. A Review on Evaluation Metrics for Data Classification Evaluations. Int J Data Min Knowl Manag Process 2015 Mar 31;5(2):01-11 [FREE Full text] [doi: 10.5121/ijdkp.2015.5201]

54. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015 Presented at: 3rd International Conference on Learning Representations; May 7-9, 2015; San Diego, CA URL: https://arxiv.org/pdf/1409.1556v6.pdf

55. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV URL: https://doi.org/10.1109/CVPR.2016.308 [doi: 10.1109/cvpr.2016.308]

56. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition; June 26-July 1, 2016; Las Vegas, NV URL: https://doi.org/10.1109/CVPR.2016.90 [doi: 10.1109/cvpr.2016.90]

57. Iandola F, Han S, Moskewicz M, Ashraf K, Dally W, Keutzer K. arXiv. URL: https://arxiv.org/abs/1602.07360 [accessed 2020-11-18]

58. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. Going deeper with convolutions. 2015 Presented at: IEEE Conference of Computer Vision and Pattern Recognition; June 7-12, 2015; Boston, MA URL: https://doi.org/10.1109/CVPR.2015.7298594 [doi: 10.1109/cvpr.2015.7298594]

59. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. Mobilenetv2: Inverted residuals and linear bottlenecks. 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT URL: https://doi.org/10.1109/CVPR.2018.00474 [doi: 10.1109/cvpr.2018.00474]

60. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. 2017 Presented at: 31st AAAI Conference on Artificial Intelligence; Feb 4-9, 2017; San Francisco, CA URL: https://dl.acm.org/doi/10.5555/3298023.3298188

61. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT URL: https://doi.org/10.1109/CVPR.2018.00716 [doi: 10.1109/cvpr.2018.00716]

62. Chollet F. Xception: Deep learning with depthwise separable convolutions. 2017 Presented at: 2017 IEEE conference on computer vision and pattern recognition; July 21-26, 2017; Honolulu, HI URL: https://doi.org/10.1109/CVPR.2017.195 [doi: 10.1109/cvpr.2017.195]

63. Raychaudhuri S. Introduction to Monte Carlo simulation. 2008 Presented at: 2008 IEEE Winter Simulation Conference; Dec 7-10, 2008; Miami, FL URL: https://doi.org/10.1109/WSC.2008.4736059 [doi: 10.1109/wsc.2008.4736059]

64. Livingston EH. Who was student and why do we care so much about his t-test? J Surg Res 2004 May 01;118(1):58-65. [doi: 10.1016/j.jss.2004.02.003] [Medline: 15093718]

65. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159 [FREE Full text] [doi: 10.1037/0033-2909.112.1.155]

66. Nakagawa S, Cuthill I. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev 2007 Nov;82(4):591-605 [FREE Full text] [doi: 10.1111/j.1469-185x.2007.00027.x]

67. Subrahmanyam M, Maheshwari R, Balasubramanian R. Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking. Signal Process 2012 Jun;92(6):1467-1479 [FREE Full text] [doi: 10.1016/j.sigpro.2011.12.005]

68. Velmurugan K, Baboo SS. Image Retrieval using Harris Corners and Histogram of Oriented Gradients. Int J Comput Appl 2011 Jun 30;24(7):6-10 [FREE Full text] [doi: 10.5120/2968-3968]

69. Nguyen D, Pham T, Baek N, Park K. Combining Deep and Handcrafted Image Features for Presentation Attack Detection in Face Recognition Systems Using Visible-Light Camera Sensors. Sensors (Basel) 2018 Feb 26;18(3):1-29 [FREE Full text] [doi: 10.3390/s18030699] [Medline: 29495417]

70. Breiman L. Random forests. Mach Learn 2001 Apr;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]

71. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV URL: https://doi.org/10.1109/CVPR.2016.319 [doi: 10.1109/cvpr.2016.319]

Abbreviations

ANN: artificial neural network
CAD: computer-aided diagnosis
CNN: convolutional neural network
GI: gastrointestinal
HOG: histogram of oriented gradients
KNN: k-nearest neighbor
LBP: local binary pattern
LSTM: long short-term memory
mAP: mean average precision
mAR: mean average recall
MLBP: multilevel local binary pattern
PCA: principal component analysis
RF: random forest
SVM: support vector machine

XSL•FO
RenderX

Original Paper

Engaging Unmotivated Smokers to Move Toward Quitting: Design of Motivational Interviewing–Based Chatbot Through Iterative Interactions

Fahad Almusharraf[1*], MASc; Jonathan Rose[1*], PhD, FIEEE, FACM, FM NAE, FRSC; Peter Selby[2,3*], MBBS, CCFP

[1]The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada

[2]Nicotine Dependence Clinic, Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada

[3]Dalla Lana School of Public Health, Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Jonathan Rose, PhD, FIEEE, FACM, FM NAE, FRSC
The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
Faculty of Applied Science & Engineering
University of Toronto
10 King's College Road
Toronto, ON, M5S 3G4
Canada
Phone: 1 4169786992
Email: jonathan.rose@ece.utoronto.ca

Abstract

Background: At any given time, most smokers in a population are ambivalent with no motivation to quit. Motivational interviewing (MI) is an evidence-based technique that aims to elicit change in ambivalent smokers. MI practitioners are scarce and expensive, and smokers are difficult to reach. Smokers are potentially reachable through the web, and if an automated chatbot could emulate an MI conversation, it could form the basis of a low-cost and scalable intervention motivating smokers to quit.

Objective: The primary goal of this study is to design, train, and test an automated MI-based chatbot capable of eliciting reflection in a conversation with cigarette smokers. This study describes the process of collecting training data to improve the chatbot's ability to generate MI-oriented responses, particularly reflections and summary statements. The secondary goal of this study is to observe the effects on participants through voluntary feedback given after completing a conversation with the chatbot.

Methods: An interdisciplinary collaboration between an MI expert and experts in computer engineering and natural language processing (NLP) co-designed the conversation and algorithms underlying the chatbot. A sample of 121 adult cigarette smokers in 11 successive groups were recruited from a web-based platform for a single-arm prospective iterative design study. The chatbot was designed to stimulate reflections on the pros and cons of smoking using MI's running head start technique. Participants were also asked to confirm the chatbot's classification of their free-form responses to measure the classification accuracy of the underlying NLP models. Each group provided responses that were used to train the chatbot for the next group.

Results: A total of 6568 responses from 121 participants in 11 successive groups over 14 weeks were received. From these responses, we were able to isolate 21 unique reasons for and against smoking and the relative frequency of each. The gradual collection of responses as inputs and smoking reasons as labels over the 11 iterations improved the F1 score of the classification within the chatbot from 0.63 in the first group to 0.82 in the final group. The mean time spent by each participant interacting with the chatbot was 21.3 (SD 14.0) min (minimum 6.4 and maximum 89.2). We also found that 34.7% (42/121) of participants enjoyed the interaction with the chatbot, and 8.3% (10/121) of participants noted explicit smoking cessation benefits from the conversation in voluntary feedback that did not solicit this explicitly.

Conclusions: Recruiting ambivalent smokers through the web is a viable method to train a chatbot to increase accuracy in reflection and summary statements, the building blocks of MI. A new set of 21 *smoking reasons* (both for and against) has been identified. Initial feedback from smokers on the experience shows promise toward using it in an intervention.

XSL•FO
RenderX

Introduction

Background

Cigarette smoking contributes to more than 6 million annual preventable deaths worldwide [1]. Canadians face the same threat, with 4.6 million Canadians smoking cigarettes in 2017 [2]. Although there are many evidence-based smoking cessation interventions, including behavioral counseling and medication [3], these are only effective in motivated smokers. The majority of smokers, however, are ambivalent about smoking and are not actively seeking help [4,5]. A key step is to motivate smokers to seek help, which has traditionally been done by clinicians and therapists during clinical encounters. Motivational interviewing (MI) [6] has been shown to be an effective approach.

MI is a patient-centered collaborative counseling method that attends to the language of change as a way to resolve the common problem of ambivalence [6]. A trained MI therapist uses 4 processes: engaging the patient, focusing on what matters to the patient to identify a potential target for behavior change, eliciting patients' motivations to change, and planning concrete steps to achieve the desired outcomes. The underlying spirits of MI are compassion, acceptance, partnership, and evocation, rather than directing the patients toward solutions. A practitioner of MI listens for preparatory *change talk* (desires, abilities, reasons, and needs to change) and uses open-ended questions, affirmations, reflections, and summary statements to elicit a commitment to change and preparation to change. The practitioner identifies steps already taken that can be built upon or amplified to achieve the desired outcome. MI has been shown to achieve a small-to-medium effect size across a variety of health behaviors, including smoking cessation [7,8].

One barrier to MI-based therapy is the time it takes for therapists to learn this modality, to gain mastery, and to be consistent in their responses over time with their clients. Besides, there are not enough health care professionals to provide such interventions in person to all smokers who might benefit from MI. If a computer-based conversational agent (also known as a chatbot) could successfully employ MI techniques, it would be easy to scale the access of many smokers to this known successful intervention. It would be both low cost and much easier to access through the many chat platforms available today. However, there are substantial barriers to overcome to create such a conversational agent.

Conversational Systems

Until recently, it has been difficult for computers to comprehend and respond appropriately to free-form text answers. This capability would allow for reflection and summary statements necessary to the key *change talk* goal of MI. Recent advances in the natural language processing (NLP) field have made it feasible to classify free-form answers into categories, which can enable a computer to select from appropriate, category-specific answers [9]. NLP focuses on the extraction and processing of specified or implied information in linguistic expressions. Conversational systems receive text or speech utterances from humans as input and generate one or more responses as output. Longer conversation chatbots (such as Microsoft's social chatbot Xiaolce [10,11]) require 3 capabilities: the ability to extract meaning out of utterances, the ability to maintain the conversation context, and the ability to generate appropriate responses [12].

Conversational systems typically use 1 of the 2 approaches to obtain information from linguistic expressions: a rule-based approach (such as the well-known 1960s-era ELIZA chatbot [13]) and a probabilistic approach [14]. In a rule-based approach, the processing is based on rules defined by expert knowledge in a specific domain. An embodiment of the rule might be the creation of a conversation *tree*, which dictates what question to ask after a specific answer is given. In contrast, the probabilistic NLP approach learns how to classify answers and possibly generate responses from a corpus of training text that illustrates many examples of related conversations.

Modern chatbots use a combination of rule-based and probabilistic approaches—the natural language understanding (NLU) models inform the rules of the conversation context and response generation. Moreover, chatbots are increasingly being used for mental health using *chat* platforms, such as Messenger (Facebook), WhatsApp (Facebook), and WeChat (Tencent) [15-17]. For example, Woebot [18,19] helps individuals with anxiety and depression using cognitive behavioral therapy and was shown to be effective in treating depression. ElizzBot is available for consulting family caregivers [20]. Although these systems are available for people already seeking help, there are few previous studies on automated chatbots that address the step of motivating individuals to seek behavior change.

Objectives

The long-term goal of this study is to create a chatbot that helps smokers move toward the decision to quit smoking. If such a chatbot is effective, it would be very simple and low cost to deploy to interact with a large set of unmotivated-to-quit smokers. This can be done through the platforms mentioned earlier or recruitment advertisements through social media channels. This study describes the first step in the creation of such an agent: a single-arm prospective study used to refine the responses of the chatbot and report on the experience of subjects on engaging the chatbot in an automated conversation about cigarette smoking.

Methods

The first step was to design the structure of a minimal conversational agent that was both automated and employed the principles of MI. This was evolved through discussion and interaction between the MI expert and clinician, and the computer engineering and NLP experts. The MI behavior change approach is to engage in a conversation that causes self-reflection with the goal of reducing smokers' ambivalence

XSL•FO

RenderX

toward quitting smoking [21]. The first key decision was to have the agent prompt subjects to articulate both the pros and cons of smoking, as they are discussed in detail in the *Conversation Structure* section. A key feature identified was the need for the chatbot to provide a nonjudgmental conversation by reflecting responses, summarizing them, and then inviting further reflection. This prototype chatbot was then tested on a sample of smokers as described in the *Recruitment of Subjects* section and iterated on to gather training data to augment the set of pro and con categories the chatbot can correctly classify. The responses and labels given by subjects over the iterations were used to train and improve the NLU classifiers employed by the chatbot.

Recruitment of Subjects

Subjects were recruited through the web from the Prolific platform [22], a website that allows researchers to offer individuals the opportunity to participate in human research experiments in exchange for financial compensation. Prolific allows researchers to select specific features from a large demographic of more than 60,000 individuals. Many of the participants reside in the United Kingdom where the company is based, but there is a significant number of participants from around the world. Textbox 1 lists the inclusion and exclusion criteria used in the subject recruitment. The *prolific rating* in Textbox 1 is the percentage of studies for which the participant has been approved in prior studies, meaning that the person running the experiment agreed that the participant's work was acceptable. It is also worth noting that there were no criteria related to a subject's motivation to quit, as this phase of the research aims to collect data from subjects with different motivations toward quitting.

Textbox 1. Inclusion criteria and exclusion criteria.

Inclusion criteria:

- English speaker
- Smokes cigarettes daily
- Resides in the United Kingdom, the United States, or Canada
- Aged between 16 and 60 years

Exclusion criterion:

- Prolific rating <9

Procedure

Subjects were presented with a consent form on the Prolific website, detailing the study's information and asking them to converse with the chatbot for the purpose of training it. With consent, subjects were then asked to visit the website, where the chatbot is deployed, and complete 2 tasks: first, to converse with the chatbot, and second, to give feedback on the overall experience of conversing with the chatbot and to make suggestions for improvement. The latter was prompted by the following question at the end of their engagement with the chatbot: "Before you finish the study, please take some time to comment on your experience chatting with the chatbot. Other than its ability to understand more of your responses, what kind of advice can you give us to improve it?"

Participants were recruited in groups of 10 or 11 participants at a time. This allowed us to retrain the chatbot in between groups and improve its classification accuracy for the next group. It also enabled the incremental addition of new categories for and against smoking, which form the central part of the conversation, as described in the *Conversation Structure* section. After each group was recruited, we determined whether the training was sufficient or if more groups were needed based on the number of new distinct categories being discovered and the accuracy of the classification achieved by the smoking reasons classifier.

Conversation Structure

The chatbot delivers the conversation to subjects in 3 stages: introduction, reflection, and ending. During the introduction stage, the chatbot describes its purpose to subjects and asks for permission to continue the conversation. The reflection stage is the core of the conversation. It asks most of the questions with the goal of eliciting reflective statements about smoking behavior. To establish this reflection, the chatbot engages subjects in 2 exercises. First, it uses the MI running head start technique [21] by asking subjects to give reasons they have for smoking (called *pros*) and reasons they have against smoking (called *cons*). We also referred to these pros and cons as *smokingreasons*. Second, it follows this with a set of questions specifically adapted to each pro or con, which stimulate subjects to contemplate the influence of each pro and con on their behavior. Figure 1 illustrates the overall flow of the key parts of the conversation in the reflection stage. The specifics shown in Figure 1 are explained in detail in the following sections.

Figure 1. Overall flow of reflection phase of chatbot.
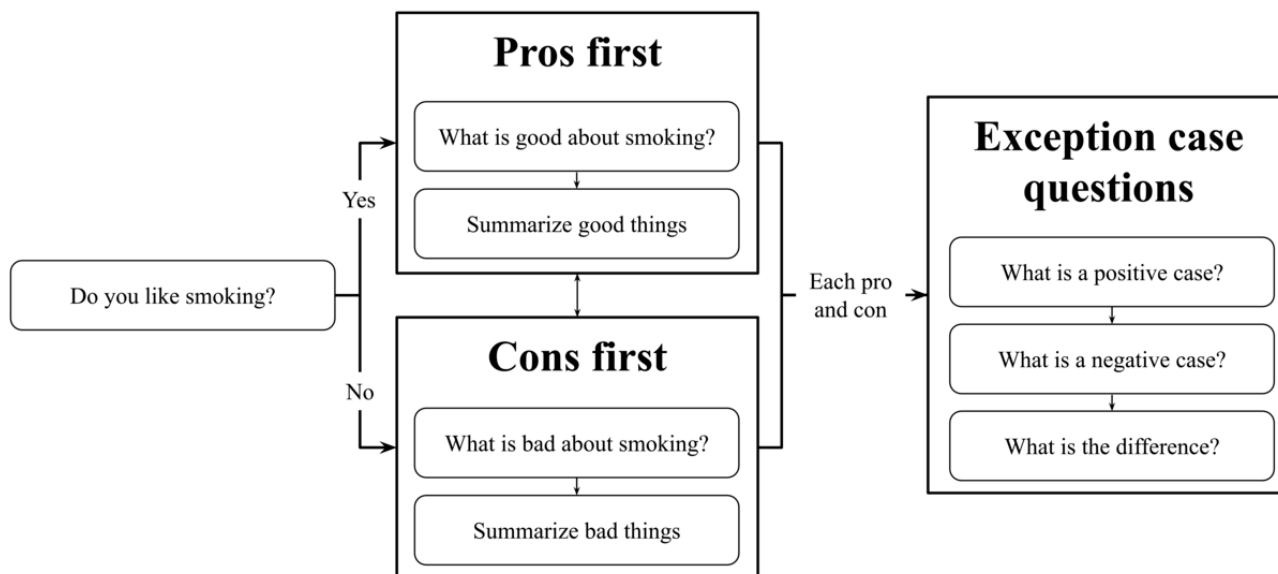
Response Generation

In general, responses are selected based on the current state of the conversation and the chatbot's classification of subjects' responses. There are 2 types of scripted responses that the chatbot can generate. The first type includes verbatim responses in expectation of a certain situation in the conversation tree. For example, the chatbot always asks subjects, "Do you like to smoke?" when it begins the reflection stage of the conversation. The second type is a scripted response where the chatbot can generate responses that contain variables taken from conversing with the subject. For example, the chatbot can generate the following variable response: "You said '...', which I believe can be classified as '...'." The first ellipsis in the response is a sentence the chatbot is recalling, and the following ellipse is its understanding of the sentence. In both types of responses, the chatbot has different variations of sentences that it can choose from to reduce repeating itself in the conversation. At any point in response generation, the chatbot chooses randomly from the set of available verbatim or variable responses.

Running Head Start

After the introduction stage of the conversation, the chatbot begins by deploying the running head start technique, which explicitly asks subjects for their pros and cons, for and against smoking. This approach provides a concrete basis for discussion, and it has been used in MI as a way to kick-start change talk, which is the eventual goal of the study [21].

Subjects are asked the following 3 questions: (1) Do you enjoy smoking? (2) What is good about smoking? and (3) What is bad about smoking? The answer to the first question determines the subject's sentiment toward smoking and is used to order the 2 subsequent questions. If the subject's response to the first question is positive, then the subject is asked, "What is good about smoking?" first. Similarly, if the response is negative, then the subject is asked, "What is bad about smoking?" first. This ordering aligns with the MI principle of keeping the conversation open ended and client centered [21]. The purpose of the second and third questions is to accumulate the subject's

pros and cons of smoking to pursue the key MI concepts of reflection and summarizing [21].

When responding to requests for the pros and cons of smoking, the subject provides free-form textual responses. NLU requires a corpus of training data that contains examples of the free-form responses together with labels that classify the response into a relevant category. A key aspect of this training phase of the chatbot is to determine the name and quantity of distinct categories of pros and cons—the smoking reasons—that would exist in all ambivalent smokers. Once these categories are established, we will be able to provide tailored responses to the subjects during an intervention study, which is discussed at the end of this paper.

The chatbot provides an initial response to each pro or con given by the subject, which reflects a classification (ie, categorization) of their pro or con. This mimics the general MI approach to communicate understanding by the MI practitioner to their clients. For example, when a subject says, "Smoking gives me a pleasurable and happy feelings," the chatbot replies with, "I understand, you get pleasure and satisfaction from smoking." This confirmatory reply reflects an apparent understanding of the chatbot. Reflection is most useful to subjects when they write long utterances that can be distilled into smaller ones or when they express multiple pros or cons that fall under the same category [21].

Once the subject has finished providing all of their pros and cons, the chatbot summarizes by repeating everything the subject has written, verbatim, within each category. For example, the chatbot might write, "You said 'the way it makes me feel', which I believe is essentially feeling good or satisfaction." In addition, the chatbot responds correctly if subjects write multiple pros or cons that are classified into the same category. It does that by summarizing each category in one utterance. For example, the chatbot summarizes 2 pros as pros in the stress category by saying, "You said 'calms nerves', and 'relieves stress', which I think are related to stress." To show how the chatbot applies the running head start technique with subjects,

Textbox 2 provides an example taken from a conversation with a recruited subject.

In summary, this first exercise of the reflection stage in the design of the conversation follows several MI principles: it aligns with the subject's sentiment toward smoking to determine whether pros or cons should be discussed first. Second, the conversation explores the pros and cons, reflecting each one in a shorter form. Finally, the chatbot brings all the pros and cons together to summarize at an appropriate point in the conversation. The *Exception Case Questions* section describes specific follow-up questions for each summarized pro and con that were given by the subject.

Textbox 2. Example of reflection and summarization in the conversation.

Conversation example:

- Bot: "Do you like to smoke?"

- Subject: "I do"

- Bot: "What do you enjoy about smoking?"

- Subject: "It is a great way to kill some time between tasks"

- Bot: "I get it, smoking is helpful because it provides a useful distraction for you"

- Bot: "What else do you enjoy about smoking?"

- Subject: "it is relaxing when doing somehthing stressful"

- Bot: "Let me see if I understand: smoking relieves you of stress"

- Bot: "You said 'It is a great way to kill some time between tasks,' which I think is related to distraction"

- Bot: "You also said 'it is relaxing when doing something stressful,' which I believe can be classified as stress"

- Bot: "Did I summarize everything back correctly?"

- Bot: "Okay. Let's now chat about the bad things about smoking"

Exception Case Questions

One of the biggest determinants that predict whether someone seeking change will take action is self-efficacy [23], which can be summarized as a person's belief in their own competence to achieve the change. To stimulate reflection toward this end, the chatbot invites the person to reflect on the reasons they gave for and against smoking. There are 2 desired outcomes to this approach: first, to have subjects enhance their own understanding of how a specific pro or con leads them to smoke or not smoke, and second, to identify situations where subjects were able to refrain from smoking and to reflect on how and why that was possible and how they might *generalize* from those situations and be able to resist the urge to smoke more often.

To achieve these outcomes, the chatbot enquires about exceptions to the subject's usual behavior. For each given pro or con, the chatbot asks 3 questions. The first question asks the subject to recall a situation in which they experienced the pro or con and how it led them to smoke (called the *positive* case). The second question asks them to remember a situation in which they experienced the same pro or con but did not end up smoking (called the *negative* case). The third question asks the subject to identify the difference between the positive and negative cases by reflecting on what might have caused the difference in outcome. These questions are designed to have the subject contemplate what triggers them to smoke and elicit examples when they successfully deferred or avoided smoking. These questions are consistent with the MI approach using a guiding form of communication rather than a directive form [21]. An example of the 3 questions taken from our training data collection is shown in Textbox 3.

Textbox 3. An example of the 3 exception case questions.

Bot: "Please describe a time when you experienced stress and ended up smoking."

Subject: "Smoking is associated with 'de-stressing' for me. If I feel anxious smoking takes the edge off."

Bot: "Please describe a time when you experienced stress and didn't end up smoking"

Subject: "Before I got divorced my ex-husband used to really stress me out but he also made me feel bad about smoking. Many times he would be stressing me out and I would not choose to smoke around him."

Bot: "Think back to the time when you were able to prevent yourself from smoking. What made it different from when you did smoke?"

Subject: "The emotional baggage involved with having that cigarette far out weighed any pleasure I would gain from smoking it, so I didn't do it."

Smoking Reason Categories and Iterative Improvement

Before the first training iteration, it was necessary to have an initial set of smoking reasons (pro and con) categories so that the NLU model can attempt to classify them to inform the chatbot how to respond. The original set of 10 reasons was taken from the following 2 sources: the first 6 reasons came from the "Why do I Smoke" questionnaire [24]: pleasure, stimulation, stress, routine, craving, or "having something in their hand." An additional 4 reasons were deemed common reasons for smoking by a smoking cessation and MI expert: addiction, health, cost, and cool.

Although these 10 categories generally indicate pro smoking reasons, some of them could also be considered as a reason against smoking. For example, a con in the stress category is smokers' feeling of anxiety when they run out of cigarettes. That is something bad about smoking. Thus, for each category, we also determined if a reason was a pro, a con, or both. This was important to detect, as the reflection and summarization responses selected by the chatbot are different depending on whether the reason is a pro or a con.

In addition to the abovementioned 10 categories, in the training study, the chatbot asks subjects to determine if the classification selected by the chatbot is correct, and if not, to give the correct category, from a list. In the case that the subject believes that a list of categories does not include the reason, the chatbot asks subjects to suggest a new category. We refer to this category confirmation process as the *training mode* of the chatbot and is illustrated with 2 examples in Textbox 4.

Textbox 4. Chatbot training mode.

Bot: "Do you have more positive things about smoking? Tell me if you can think of any"

Subject: "Helps me to bond with other smokers and socialize"

Bot: "I understood 'Helps me to bond with other smokers and socialize' as something related to 'Socializing'"

Bot: "Am I right?"

Subject: "Right"

Bot: "Thank you for confirming my knowledge"

Bot: "I understand, smoking is a social habit for you"

Bot: "What else do you find positive about smoking?"

Subject: "I like the way the smoke tastes"

Bot: "I understood 'I like the way the smoke tastes' as something related to 'Feeling good'"

Bot: "Am I right?"

Subject: "Wrong"

Bot: "Is it on of the following options? Addiction, Stimulation, ... , Other"

Subject: "Other"

Bot: "Can you please tell me what it is in one or two words?"

Subject: "Flavor"

Bot: "Thank you for teaching me something new!"

Bot: "What else do you dislike about smoking?"

Subject: "yellow fingers"

Bot: "I understood 'yellow fingers' as something related to 'Health'"

Bot: "Am I right?"

Subject: "Wrong"

Bot: "Is it on of the following options? Addiction, Stimulation, ... Physical Appearance, ..., Other"

Subject: "Physical Appearance"

After each group of subjects finished their respective conversation, new information about the smoking reason categories emerged and required consideration by the researchers. This information can be one of the following, as shown in Textbox 4: (1) confirmation of category detection, (2) correction of category detection, and (3) introduction of a new category. To ensure the validity and consistency of the category information provided by subjects, we reviewed each confirmation, correction, and introduction after each training group and before retraining the chatbot for the next one. In our review, we validate that confirmations and corrections are indeed valid and are not the results of typing or subject error.

We also considered the introduction of new categories by reviewing the conversation and subject responses that suggested them. If, on discussion, we could not reach consensus that a response fit into an existing category, then a new category was

created and given a descriptive name. As a result, we need to create scripted responses so that the chatbot can reflect and summarize the new categories on detection. The process for scripting new responses is similar to how we agree on a new category; it is decided based on discussion among the researchers. As we follow an iterative strategy in engaging subjects and training the chatbot, more pro and con categories appeared, as subjects introduced new reasons to the chatbot through this training mode in the conversation.

Analysis

The primary goal of the study is to build a data set of example pro and con reasons for smoking and their categories. The secondary goal is to make the conversation experience with the chatbot as pleasant as possible. To evaluate these goals, we counted the number of examples of the pros and cons and their categories. We also calculated the precision, recall, and F1 score of the chatbot's classifiers on a set of examples that it has not been trained on to measure its performance. Precision measures the percentage of correct detections from all the detections a classifier makes on the test data, whereas recall measures how often the classifier incorrectly misclassified responses or was unable to determine any classification. The F1 score is the harmonic mean of precision and recall, and it measures the performance of the classifier in its ability to generate the correct detections (precision) and not miss any of them (recall). Table 1 lists the 3 measurements used for evaluating the 2 classifiers. All 3 metrics are calculated as micrometrics: true positive (TP), false positive (FP), and false negative (FN) are calculated globally across the classifier classes.

Table 1. Definitions of precision, recall, and F1 score.

Measurement	Definition
Precision	TP^a (TP + FP^b)
Recall	TP / (TP + FN^c)
F1 score	TP / (TP + 0.5 x [FP + FN])

[a]TP: true positive.

[b]FP: false positive.

[c]FN: false negative.

When calculating the metrics in Table 1, a TP event is defined as the event when the chatbot can correctly detect the presence of a category in the subject's utterance and generate an appropriate response. An FP event is observed when the chatbot generates an incorrect classification and provides an off-target response to a subject's utterance. Finally, an FN is observed when the chatbot is unable to generate any classification on a given subject's utterance.

Ethics Review

The University of Toronto Health Science Research Ethics Board (REB) reviewed and approved the study. The REB protocol number is 35962 and was approved on May 28, 2018.

Results

Participants

A total of 121 participants completed the study in 14 weeks. Table 2 gives the participants' demographic information, including their age, sex, smoking frequency, last quit attempt, employment status, and country of residence. This information was entered by participants when they first registered with Prolific, and it is possible that not all fields have responses from all participants. Information that was not provided by participants to Prolific is marked as *missing* in Table 2.

Table 2. Demographics of the subjects in the study (N=121).

Characteristics	Values
Age (years), mean (SD)	35.2 (9.8)
Age (years), n (%)	
16-19	0 (0.0)
20-29	40 (32.8)
30-39	42 (34.4)
40-49	23 (18.9)
50-59	13 (10.7)
60	1 (0.8)
Missing	3 (2.5)
Sex, n (%)	
Female	60 (49.2)
Male	59 (48.4)
Missing	3 (2.5)
Smoking frequency, n (%)	
Once a day	5 (4.1)
2-5 times a day	20 (16.4)
6-10 times a day	29 (23.8)
11-19 times a day	48 (39.3)
≥20 times a day	19 (15.6)
Missing	1 (0.8)
Last quit attempt, n (%)	
Never	16 (13.1)
>12 months	30 (24.6)
7-12 months	10 (8.2)
4-6 months	14 (11.5)
1-3 months	15 (12.3)
Currently trying	14 (11.5)
Missing	23 (18.8)
Employment status, n (%)	
Full time	61 (50.0)
Part time	21 (17.2)
Not in paid work	17 (13.9)
Unemployed (and job seeking)	15 (12.3)
Other	7 (5.7)
Missing	1 (0.8)
Country of residence, n (%)	
United Kingdom	66 (54.1)
United States	50 (41.0)
Canada	6 (4.9)

Participant Interaction

The chatbot received 6568 responses from the 121 participants, where 4271 were free-form responses and 2297 were selected from the suggested responses from the chatbot. The mean time spent by each participant interacting with the chatbot was 21.3 (SD 14.0) min. The shortest conversation was 6.4 min long, whereas the longest conversation lasted for 89.2 min. The longest conversation was a result of a specific participant having more pros and cons about smoking as well as crafting long responses to the chatbot.

Smoking Reasons Data Set

An additional 11 distinct categories of smoking reasons separate from the original 10 categories were identified. Table 3 provides the entire list and indicates whether a category can be considered as a pro, con, or both. It also provides an example statement from a study participant in each category.

This study produced a data set of 1010 samples of how participants expressed 21 distinct categories of why smoking is good or bad. Of these 1010 samples, 79 samples and 10 categories were synthesized by the researchers before the initial group of participants, as described earlier. The remaining 931 samples and 11 categories were generated through the input of the recruited participants. Of the 1010 samples, 490 expressed pros about smoking and 520 expressed cons about smoking. Of the 21 distinct categories, 5 are pro categories (boredom, cool, feel good, something in my hand, and stimulation), 6 are con (cost, dirty, fire hazard, physical appearance, poor role model, and stigma), and the remaining 10 are both pro and con (addiction, distraction, flavor, focus, health, routine, smell, smoking restriction, social, and stress). Table 3 also lists the number of responses acquired in each reason category.

Table 3. A list of all the smoking reason categories used in the conversation.

Category and pro or con	Count	Example
Addiction		
Con	1	"The first thing I think of when I wake up is my first cigarette"
Pro	55	"I crave nicotine"
Boredom		
Pro	16	"I smoke out of boredom"
Cool		
Pro	9	"Makes me look cool"
Cost		
Con	67	"Cost so much for such little joy"
Dirty		
Con	16	"The cigarette ash drops about"
Distraction		
Pro	52	"It gives me time to myself and time to think"
Con	11	"how it interrupts your work"
Feel good		
Pro	85	"I enjoy the feeling that it gives me"
Fire hazard		
Con	7	"It burns my home and furniture"
Flavor		
Pro	25	"I like the way the smoke tastes"
Con	7	"The taste smoking leaves in your mouth"
Focus		
Pro	15	"Helps me concentrate doing computer work"
Con	3	"Can't concentrate if I need a cigarette"
Health		
Pro	25	"Cigarettes help with bowel movements"
Con	133	"All the health problems smoking is linked to causing"
In my hand		
Pro	30	"Gives me something to occupy my hands with"
Physical appearance		
Con	31	"Aging or appearance change is always a fear"
Poor role model		
Con	14	"I feel guilty because my son doesn't like me smoking and nags"
Routine		
Pro	9	"It something to look forward to doing"
Con	5	"routine cigarettes"
Smell		
Pro	9	"The smell of cigarette smoke is nice"
Con	74	"Leave you smelling on fingers breath and clothes"
Smoking restriction		
Pro	2	"Sometimes smoking restrictions will force me to go outside and then I realize it's a beautiful night and i'm glad it forced me to go outside"

Category and pro or con	Count	Example
Con	22	"You can't smoke in a lot of places"
Social		
Pro	58	"I like to socialise with other people who smoke"
Con	11	"people around me do not smoke, only I do"
Stigma		
Con	62	"Being made to feel unwelcome by non smokers"
Stimulation		
Pro	37	"The sensation of the tobacco as it catches my throat"
Stress		
Pro	177	"Helps me relax and decrease stress"
Con	2	"Feeling anxious when I've run out"

Classifier Training Result

In implementing the software for the chatbot, we originally used a third-party web-based system for classification called Wit.ai [25], referred to as classifier SR1 (smoking reasons [classifier] 1). Although it made the ramp-up of the classification easier, we did not have the ability to understand and control the behavior of the *black box* classifier. This led us to build our own classifier using the NLP framework from spaCy [26], referred to as classifier SR2 (smoking reasons [classifier] 2). Using the entire data set obtained from the 121 participants (810/1010, 81.2% used as training data and 200/1010, 19.8% used as test data), we measured the performance of the 2 classifiers. Table 4 gives the overall capability of the 2 classifiers used after all the training data were collected. Although classifier SR1 has slightly better precision than SR2, the latter has a much better recall and thus gives a much better overall F1 score. Recall can informally be thought of as the chatbot's ability to correctly reflect what subjects are expressing. Higher degrees of recall means that the chatbot is able to detect more of the subjects' pros and cons and generate reflections on them.

Table 4. Precision, recall, and F1 score for the smoking reasons classifier 1 and smoking reasons classifier 2 natural language understanding classifiers.

Measurement	Smoking reasons classifier 1	Smoking reasons classifier 2
Precision	0.98	0.87
Recall	0.28	0.84
F1 score	0.44	0.86

A measure of the progress of the classifiers over the training group iterations is given in Table 5, which shows the precision, recall, and F1 score that each participant group experienced during the actual training session. (This is quite distinct from the results in Table 4 because those are given when the classifier is trained and evaluated using the entire corpus of utterances and labels.) Several things are changing as these results were acquired over time—the number of categories is increasing and the amount of training data available to train the classifier is increasing. In addition, as described, the classifier used from groups 1 to 10 was SR1, whereas group 11 used classifier SR2. Table 5 illustrates the progress of the classifier used during training, ending with a significantly better F1 score overall there as well, with classifier SR2.

XSL·FO

RenderX

Table 5. Precision, recall, and F1 score of the chatbot's pro and con detection.

Group	Precision	Recall	F1 score	Classifier used
1	0.93	0.48	0.63	SR1[a]
2	0.97	0.56	0.71	SR1
3	0.96	0.26	0.41	SR1
4	0.96	0.46	0.62	SR1
5	0.93	0.67	0.78	SR1
6	0.92	0.63	0.75	SR1
7	1.00	0.61	0.76	SR1
8	0.96	0.68	0.80	SR1
9	0.93	0.58	0.71	SR1
10	0.95	0.60	0.73	SR1
11	0.91	0.75	0.82	SR2[b]

[a]SR1: smoking reasons classifier 1.

[b]SR2: smoking reasons classifier 2.

Voluntary Free-Form Feedback

Participants were asked to voluntarily answer the following question at the end of their engagement with the chatbot: "Before you finish the study, please take some time to comment on your experience chatting with the chatbot. Other than its ability to understand more of your responses, what kind of advice can you give us to improve it?" The majority of answers to this question suggested improvements to the chatbot. However, there were other signals in the feedback that emerged. The other responses included that participants enjoyed conversing with the chatbot, found it beneficial, or were frustrated by it. To measure these data quantitatively and based on discussion among the researchers, the following 4 mutually inclusive (ie, overlapping) labels were added to each participant's feedback: improvement suggestions, enjoyment, benefit, and frustration. Textbox 5 describes each of the 4 labels. In the study, 76.9% (93/121) of participants answered the feedback questions. The following percentages of the 4 labels were observed in their feedback: 44.6% (54/121) expressed improvement suggestions, 34.7% (42/121) expressed enjoyment, 8.3% (10/121) indicated benefit, and 2.5% (3/121) expressed frustration. Textbox 6 provides some examples of the benefit and frustration feedback received from the participants in the study.

Textbox 5. Description of feedback labels.

Improvement suggestions

- Any feedback that suggested bug fixes, new capabilities, or comments relating to functionality

Enjoyment

- Feedback that indicates a positive, pleasant experience with the chatbot

Benefit

- Feedback explicitly indicating that the chatbot caused participants to reflect on their behavior or motivated them to quit

Frustration

- Any feedback that indicates the subject had a negative experience or caused a negative effect

Textbox 6. Feedback samples from the study.

Benefit

- "The study actually made me think more about quitting smoking"

Benefit

- "That was actually really very helpful. It was getting my thoughts out. I have been smoking for a long time and this was the first time believe it or not that I actually got some insight to my behavior, LOVED IT."

Frustration

- "This chatbot really needs some redesign. Punctuation seems to throw it off 'Yes' is accepted but 'Yes.' is not. Also, it's making assumptions of people which is going to make them combative, like me, when it just goes 'obviously you aren't able to stop yourself ever'."

Frustration

- "Frustrating, the questions made little sense. I had 'it smells bad' and it asked me to describe a time when the action 'it smells bad' caused you to smoke. A lot of self reflections kind of felt pointless as well. I smoke a pack a day, often without thinking about it., so pinpointing a time when something caused me to smoke is really hard."

The free-form feedback from participants gives an informative view of their experience: almost 35% (42/121, 34.7%) of participants found conversing with the chatbot a pleasant experience. In addition, more subjects expressed benefit (42/121, 8.3%) than frustration (3/121, 2.5%). We deem the frustration effects as not harmful in the long term because the target audience are unmotivated subjects, and continuing to smoke does not have a short-term harmful effect.

Discussion

Principal Findings

Iteratively recruiting participants to collect training data as well as engaging them in a conversation about smoking enabled the training and validation of the MI-based chatbot. Although our chatbot does not strictly follow the 4 processes of MI mentioned in the *Introduction* section, it does follow the spirited principles of MI. It keeps the conversation client centric by tailoring the running head start technique to subjects' sentiment toward smoking. It provides reflection and summarization to subjects' expressed pros and cons of smoking. Finally, it uses the exception case questions to each pro and con, possibly revealing situations that might resolve ambivalence. In addition, we were able to improve the performance of our chatbot and improve its ability to hold a conversation with a relatively small number of participants. Interesting information emerged from the voluntary feedback question asked at the end of the experiment. More than 1 in 3 participants enjoyed conversing with the chatbot, and the qualitative data indicated that although a small subset of participants found benefit in engaging with the chatbot, more refinements are necessary to minimize the frustration that could affect engagement in behavior change.

In the focus area of this study, there has been some prior work on smoking cessation–based chatbots. Perski et al [27] explored adding a chatbot to a preexisting smoking cessation support mobile app with a randomized control trial. Their intervention used several of the techniques we employed in this study, including identifying reasons for wanting and not wanting to stop smoking and tailoring the interactions appropriately. They showed that using a chatbot in their mobile app increased engagement and the odds of quitting success. However, they did not make use of reflective listening, which is one of the core capabilities of our chatbot.

A key part of our designed conversation uses reflective listening in the elicitation of the pros and cons of smoking and their subsequent short-form reflection and restatement. Although our primary goal was to collect training data and test our chatbot on recruited participants, we have gained insights into how to improve our chatbot for future interventions. Our experience showed that subjects generally do not articulate short descriptions of the pros and cons of smoking. Rather, they explain feelings associated with a given pro or con. For example, a subject might say, "Helps me bond with other smokers" to communicate *socializing* as a pro for smoking. Therefore, the system must have a reasonably complete set of categories for expressing the pros and cons of smoking, such as *socializing*. The name of those categories must be reflective of the many ways that subjects may express a pro or con so that a subject can agree (or disagree) correctly when the chatbot makes a classification during the training study. In addition, the reflection responses given by the chatbot are tuned to each category; thus, choosing the right number of distinct categories, so that these responses are effective, is important.

It is worth noting that in the context of this chatbot, precision is not as important as recall. High recall is desired for providing reflections to subjects. The current implementation of the NLU classifier for detecting the pros and cons is a mutually exclusive classifier. Any given utterance expressed by the subject for a pro or con for smoking only maps to one category. However, the categories of the pros and cons do partially overlap and are subjective. For example, our chatbot classifies the utterance "how it interrupts your work" into the *focus* category instead of the *distraction* category. As a result of this classification, the chatbot will respond with "I understand, smoking makes you lose focus" instead of "I get it, smoking can be a distraction for you." In this case, the generated response from the chatbot will most likely satisfy the subject, as it would for other categories that are close in meaning.

According to Pereira and Diaz [28], most health-related chatbots focus on neurological and nutritional disorders whereas smoking, which falls under the category of addiction disorders, is one of the less focused-on health problems tackled by chatbots. In addition, *consumability* (a description of users' end-to-end experience with technology) and personalization are the 2 main enablers of chatbots in the health space. In this research, we also focus on these enablers to bring chatbots to the underserviced health problem of cigarette smoking addiction. Our chatbot achieves personalization by tailoring the conversation around subjects' perspectives about the positive and negative aspects of smoking during optimization for a human-like conversation experience.

Finally, it was promising to observe that some participants gave extensive answers to the questions posed by the chatbot, and these answers were thoughtful and reflective. This suggests that when deployed as an intervention, there may be a good effect for some future participants.

Limitations and Future Improvements

This study has 4 main limitations. The first limitation is using the running head start technique. Although this technique is recommended when there is an observed absence of *change talk* [21], we are using it to create a simple basis for discussion and reflection. We recognize that for subjects who are already exhibiting change behavior, this might not be appropriate or helpful, and in the future, we plan to improve the chatbot to detect change behavior early in the conversation and employ a different strategy.

The second limitation is the assumption that subjects are unmotivated to quit; however, we know from the demographics in Table 1 that some of the subjects may be currently trying or recently tried quitting. As the chatbot uses the running head start MI strategy to elicit reflection in subjects, this class of subject may find the elicitation of pros and cons frustrating because they are already motivated to quit [21]. One possible solution to this problem is to screen motivated smokers using techniques such as the readiness ruler and ending the conversation with them. The chatbot could also lead them to

other resources appropriate for dealing with their stage of behavior change.

The third limitation is the subjects' self-reported demographic data on the Prolific platform [22]. The collected demographics in Table 1 might be outdated. One important example is the "When was your last quit attempt" information, which is reported once subjects join the platform. This information can be stale and not indicative of the current state of the subject. In our future work, we intend to point subjects to a survey they answer before entering a conversation with the chatbot. In this survey, we can ask subjects for more recent answers about variables that might have changed from when they joined Prolific.

The fourth limitation is that the chatbot finishes the conversation after receiving the response to the exception case questions. This lack of follow-on to the exception case questions or elsewhere in the conversation can frustrate subjects and possibly lead to negative unintended effects. Generating responses on general situational reflections in the exception case questions stage of the conversation requires general NLU response generation capabilities that are being actively researched. However, in the future, we hope to provide useful responses to these reflections and to continue the conversation productively.

Conclusions

This study has described the design and training of a conversational agent whose purpose is to interact with ambivalent smokers to move them toward quitting smoking. The agent employs strategies from MI and makes use of the running head start technique to launch a concrete discussion. A key aspect of the design is to allow free-form text responses to questions and the use of NLU techniques to correctly categorize the free-form responses. We were able to show a method to train the NLU engine to accurately identify responses, which is then used to select an appropriate sequence of responses. A side effect of the training was to identify 21 distinct categories of reasons for and against smoking that the training subjects helped to define. The next step is to conduct a feasibility study of the now-trained intervention and to iterate on the design again to improve its effectiveness.

Conflicts of Interest

None declared.

References

1. Reitsma MB, Fullman N, Ng M, Salama JS, Abajobir A, Abate KH, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet 2017 May;389(10082):1885-1906. [doi: 10.1016/S0140-6736(17)30819-X]
2. Canadian Tobacco, Alcohol and Drugs Survey (CTADS): Summary of Results for 2017. Health Canada. 2017. URL: https://www.canada.ca/en/health-canada/services/canadian-tobacco-alcohol-drugs-survey/2017-summary.html#n2 [accessed 2020-03-19]
3. Heydari G, Masjedi M, Ahmady AE, Leischow SJ, Lando HA, Shadmehr MB, et al. A comparative study on tobacco cessation methods: a quantitative systematic review. Int J Prev Med 2014 Jun;5(6):673-678 [FREE Full text] [Medline: 25013685]

4. Hardcastle SJ, Hancox J, Hattar A, Maxwell-Smith C, Thøgersen-Ntoumani C, Hagger MS. Motivating the unmotivated: how can health behavior be changed in those unwilling to change? Front Psychol 2015;6:835 [FREE Full text] [doi: 10.3389/fpsyg.2015.00835] [Medline: 26136716]

5. Wewers ME, Stillman FA, Hartman AM, Shopland DR. Distribution of daily smokers by stage of change: current population survey results. Prev Med 2003 Jun;36(6):710-720. [doi: 10.1016/s0091-7435(03)00044-6] [Medline: 12744915]

6. Hettema JE, Hendricks PS. Motivational interviewing for smoking cessation: a meta-analytic review. J Consult Clin Psychol 2010 Dec;78(6):868-884. [doi: 10.1037/a0021498] [Medline: 21114344]

7. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]

8. Lundahl BW, Kunz C, Brownell C, Tollefson D, Burke BL. A meta-analysis of motivational interviewing: twenty-five years of empirical studies. Res Soc Work Pract 2010 Jan 11;20(2):137-160. [doi: 10.1177/1049731509347850]

9. Jurafsky D, Martin JH. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, NJ, USA: Prentice Hall; 2021.

10. Microsoft XiaoIce. URL: https://www.msxiaobing.com [accessed 2020-01-21]

11. Like a Phone Call: Xiaoice, Microsoft's Social Chatbot in China, Makes Breakthrough in Natural Conversation. Microsoft News. 2019. URL: https://news.microsoft.com/apac/features/like-a-phone-call-xiaoice-microsofts-social-chatbot-in-china-makes-breakthrough-in-natural-conversation/ [accessed 2020-01-21]

12. McTear M, Callejas Z, Barres DG. The Conversational Interface: Talking to Smart Devices. New York, USA: Springer International Publishing; 2016.

13. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]

14. Manning CD, Schütze H. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press; 1999.

15. Facebook Messenger. URL: https://www.messenger.com/ [accessed 2020-01-21]

16. WhatsApp. URL: https://www.whatsapp.com [accessed 2020-01-21]

17. WeChat. URL: https://www.wechat.com/en [accessed 2020-01-21]

18. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WoeBot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

19. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]

20. ElizzBot. 2017. URL: https://elizz.com/elizzbot/ [accessed 2020-01-21]

21. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. New York, USA: Guilford press; 2012.

22. Prolific: Online Participant Recruitment for Surveys and Market Research. URL: https://www.prolific.co/ [accessed 2020-01-21]

23. Bandura A. Self-Efficacy: The Exercise of Control. New York, USA: W H Freeman; 1997.

24. Russell MA, Peto J, Patel UA. The classification of smoking by factorial structure of motives. J R Stat Soc Series B 1974;137(3):313. [doi: 10.2307/2344953]

25. Wit.Ai. URL: https://wit.ai/ [accessed 2020-01-21]

26. spaCy: Industrial-Strength Natural Language Processing in Python. URL: https://spacy.io/ [accessed 2020-01-21]

27. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676. [doi: 10.1177/2055207619880676] [Medline: 31620306]

28. Pereira J, Díaz O. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 4;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]

Abbreviations

FN: false negative
FP: false positive
MI: motivational interviewing
NLP: natural language processing
NLU: natural language understanding
REB: Research Ethics Board
SR1: smoking reasons classifier 1
SR2: smoking reasons classifier 2
TP: true positive

XSL•FO

RenderX

Original Paper

My Diabetes Coach, a Mobile App–Based Interactive Conversational Agent to Support Type 2 Diabetes Self-Management: Randomized Effectiveness-Implementation Trial

Enying Gong[1], MSc; Shaira Baptista[1,2], GradDipPsychSc, BSc (Hons); Anthony Russell[3,4], PhD; Paul Scuffham[5], PhD; Michaela Riddell[6], PhD; Jane Speight[1,2,7], PhD; Dominique Bird[4], MD; Emily Williams[8], PhD; Mojtaba Lotfaliany[1], MD, PhD; Brian Oldenburg[1], PhD

[1]School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[2]The Australia Center for Behavioral Research in Diabetes, Diabetes Victoria, Melbourne, Australia

[3]Princess Alexandra Hospital,, Queensland, Australia

[4]Centre for Health Services Research, The University of Queensland, Brisbane, Australia

[5]Menzies Health Institute Queensland, Griffith University, Queensland, Australia

[6]Kirby Institute, University of New South Wales, Sydney, Australia

[7]School of Psychology, Deakin University, Geelong, Australia

[8]Faculty of Health and Medical Sciences, University of Surrey, Surrey, United Kingdom

Corresponding Author:
Enying Gong, MSc
School of Population and Global Health
The University of Melbourne
333 Exhibition
Melbourne
Australia
Phone: 61 0452389420
Email: egong@student.unimelb.edu.au

Abstract

Background: Delivering self-management support to people with type 2 diabetes mellitus is essential to reduce the health system burden and to empower people with the skills, knowledge, and confidence needed to take an active role in managing their own health.

Objective: This study aims to evaluate the adoption, use, and effectiveness of the My Diabetes Coach (MDC) program, an app-based interactive embodied conversational agent, *Laura*, designed to support diabetes self-management in the home setting over 12 months.

Methods: This randomized controlled trial evaluated both the implementation and effectiveness of the MDC program. Adults with type 2 diabetes in Australia were recruited and randomized to the intervention arm (MDC) or the control arm (usual care). Program use was tracked over 12 months. Coprimary outcomes included changes in glycated hemoglobin (HbA_{1c}) and health-related quality of life (HRQoL). Data were assessed at baseline and at 6 and 12 months, and analyzed using linear mixed-effects regression models.

Results: A total of 187 adults with type 2 diabetes (mean 57 years, SD 10 years; 41.7% women) were recruited and randomly allocated to the intervention (n=93) and control (n=94) arms. MDC program users (92/93 participants) completed 1942 chats with *Laura*, averaging 243 min (SD 212) per person over 12 months. Compared with baseline, the mean estimated HbA_{1c} decreased in both arms at 12 months (intervention: 0.33% and control: 0.20%), but the net differences between the two arms in change of HbA_{1c} (–0.04%, 95% CI –0.45 to 0.36; P=.83) was not statistically significant. At 12 months, HRQoL utility scores improved in the intervention arm, compared with the control arm (between-arm difference: 0.04, 95% CI 0.00 to 0.07; P=.04).

Conclusions: The MDC program was successfully adopted and used by individuals with type 2 diabetes and significantly improved the users' HRQoL. These findings suggest the potential for wider implementation of technology-enabled

XSL•FO
RenderX

conversation-based programs for supporting diabetes self-management. Future studies should focus on strategies to maintain program usage and HbA$_{1c}$ improvement.

Introduction

Type 2 diabetes mellitus (T2DM) is a common chronic condition that places a significant burden on individuals and the health care system. The prevalence of diabetes has risen substantially over the past two decades worldwide [1]. However, clinical outcomes have not improved significantly despite considerable investment and advances in treatments and technologies over this period [2]. Delivering self-management support to people with T2DM is essential to reduce the health system burden and to empower people with the skills, knowledge, and confidence needed to take an active role in managing their own health [3].

Health coaching programs, incorporating continuous feedback and reinforcement [4] delivered to people with diabetes by health care professionals, have been shown to be an effective strategy to improve glycemic management [3,5]. However, the human delivery of such programs on a large scale is very challenging, given the health system and workforce constraints. The emergence of information and communication technologies in recent years offers the potential to deliver such programs at scale and with relatively low costs [6]. Our previous studies and others have demonstrated the effectiveness and cost-effectiveness of using technology-enabled programs, including computer-based or telephone-based coaching, in promoting behavior change and self-management among adults with T2DM [7-9]. The increasing ubiquity of smartphones and other smart devices now offers the potential to further optimize the delivery of such programs at scale by the use of apps.

Despite the increasing use of mobile apps for health purposes, reviews have found that existing digital health solutions are not generally able to meet the needs of people with diabetes, and more evidence is required before their wider scale-up [10,11]. Most commercially available apps employ limited use of behavior change techniques and inadequate features, other than self-monitoring [12,13]. Furthermore, people with diabetes expect apps to be engaging; incorporating multiple functions; and covering a broad range of content, including psychological and emotional support [14]. Although some recent systematic reviews and meta-analyses have indicated a modest effect for app-based interventions to support diabetes self-management [10,11,15], there still remains a great deal of uncertainty about their feasibility, acceptability, and effectiveness in the *real world* [10,15-17].

By adapting our team's previous effective and cost-effective Telephone-Linked Care for Diabetes (TLC diabetes) program [8] to a more contemporary technology platform, we designed the My Diabetes Coach (MDC) program, which incorporates an app-based interactive embodied conversational agent, *Laura*, to provide more accessible and engaging self-management support, monitoring, and coaching to adults with T2DM in Australia. We undertook a hybrid effectiveness-implementation study [18] with a randomized controlled design. The study aims to evaluate both the implementation of the MDC program over 12 months and its effectiveness. In this paper, we report on the program adoption and use during the trial and program effectiveness in terms of the coprimary outcomes and selected secondary outcomes. The report of the study is in line with the CONSORT (Consolidated Standards of Reporting Trials) guideline and the extension guideline of the CONSORT-eHealth guidelines.

Methods

Study Design Overview

This trial is a two-arm, open-label, randomized controlled trial with participants recruited between June 2016 and April 2017 in Australia. The trial was registered before recruitment (Australia New Zealand Clinical Trials Registry ID: ACTRN12614001229662). Full ethics approval was granted by the University of Melbourne's Human Research Ethics Committee (Ethics ID 1442433). Participants provided written informed consent and returned the informed consent forms to the research team via email or fax, including permission for their general practitioners (GPs) to regularly share clinical data with the study team.

Study Participants

Adults (aged ≥18 years) diagnosed with T2DM, registered with the National Diabetes Service Scheme (NDSS) for less than 10 years, with basic English language skills, who had access to an internet-enabled smart device with an up-to-date operating system (at least iOS 8.0 for Apple and 4.2 for Android) were eligible to participate in the study. Participants were ineligible if they were pregnant or planning to become pregnant, had severe comorbid conditions that would compromise their participation, or did not have stable doses of diabetes-related medication over the previous 4 weeks or more.

To assist with recruitment, on behalf of the research team, the NDSS sent invitation letters to registered adults with T2DM living in the Australian states of Queensland, Victoria, and Western Australia. These 3 states comprise 54.5% of all Australians with diagnosed diabetes [19]. Recruitment information was also posted on social media via the University of Melbourne website, Facebook, and Twitter and was made available via Diabetes Australia and state-based diabetes

organizations (eg, newsletters or magazines). This approach enabled people from all Australian states or territories to enter the study. People who submitted an expression of interest were contacted by the research team by telephone within 48 hours to complete eligibility screening. Eligible participants who provided consent and completed the baseline survey were formally enrolled in the study.

Randomization

Participants were randomly allocated to the intervention or control arm using a 2x4 block randomization sequence, programmed into a Redcap data management system [20], with the individual as the unit of randomization. The program coordinator informed participants and their GPs of the randomization allocation via telephone or email. Due to the nature of the intervention, participants, their GPs, and the program coordinator could not be blinded to participant study allocation. A statistician who verified the data analysis was blinded to the randomization.

Intervention and Control

Intervention: MDC Program

Participants allocated to the intervention arm received access to the MDC program for up to 12 months. The overall program comprises 5 components: the MDC app; a printed user guide; the MDC website; an optional blood glucose meter with Bluetooth capability (Accu-Chek Aviva Connect, Roche Diabetes Care); and a small number of brief, structured interactions with a program coordinator, primarily for technical assistance. The MDC app was adapted from the previous TLC diabetes program [8] and was co-designed with the support of the Bupa Foundation and a technology company named Clevertar. Multimedia Appendix 1 details the MDC program and these 5 components.

MDC delivers personalized support, monitoring, and motivational coaching via an embodied conversational agent, *Laura*, through a series of modules covering blood glucose monitoring, healthy eating, physical activity, medication taking, and foot care. The conversational scripts and algorithms guiding each individual's progress were designed by applying behavior change theories and techniques, including the transtheoretical model [21], social cognitive theory [22], gamification [23], and other concepts derived from chronic disease self-management [24]. Algorithms were further tailored according to the clinical targets and recommendations provided by each participant's GP.

Participants were encouraged to use the app weekly to complete online modules by chatting with *Laura* or touching buttons on the screen. Each *appointment* module with *Laura* began with a review of progress with feedback, education and counseling on the chosen topic, and incorporated tips on overcoming barriers, followed by a short quiz and closing remarks.

In addition to the MDC app, participants were also encouraged to regularly access the user guide and the MDC website and to join the discussion forums on diabetes self-management topics posted fortnightly by the program coordinator on the website. The program coordinator, supported by a web-based user management portal, assisted participants in dealing with system-generated technical alerts by communicating with participants, their GPs, and Clevertar. The program coordinator also telephoned participants after 1, 4, 8, 12, and 24 weeks of program access to answer questions, to troubleshoot technical issues, and to encourage program use.

Control Arm

Participants in the control arm were encouraged to continue their routine diabetes self-care, including access to health care services, resources accessed via NDSS, and the diabetes not-for-profit organizations in their states. They received a quarterly project newsletter to maintain their interest in the study. Following the 12-month data collection, participants in the control arm received access to the MDC program if they wished.

Measurement and Outcomes

We used the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework to evaluate the impact of the MDC intervention, which covered the 5 dimensions in terms of reach, effectiveness, adoption, implementation, and maintenance [25]. This paper focuses on program adoption, use, and effectiveness.

Program Adoption and Use

Program adoption and use were tracked using the program management portal. The key metrics included the number of completed chats with *Laura*, the total duration of chats, and the number of blood glucose levels uploaded by participants. Other metrics included the number of technical and clinical alerts generated by the system for the project coordinator and the number of posts on the web-based discussion forum of the MDC website.

Program Effectiveness

Program effectiveness was measured by both clinical and psycho-behavioral outcomes. The coprimary outcomes were changes (12 months compared with baseline) in glycated hemoglobin (HbA_{1c}) and health-related quality of life (HRQoL), which were examined in terms of between-arm differences. The secondary time point of analysis examined the change between baseline and 6 month. HbA_{1c} (reported as % and mmol/mol) was measured through a pathology blood test that each participant's GP requested. HRQoL was assessed via participants' completion of the Assessment of Quality of Life (AQoL)-8D scale, which is a 35-item multi-attribute utility instrument covering 8 dimensions focused on independent living, happiness, mental health, coping, relationships, self-worth, pain, and senses [26,27]. The AQoL-8D provides a utility score ranging from –0.04 to 1, where higher scores indicate better utility value in HRQoL.

Selected secondary outcomes reported in this paper include anxiety and depressive symptoms, assessed by the Hospital Anxiety and Depression Scale (HADS) [28]; diabetes-specific distress, assessed by the Problem Areas in Diabetes scale [29]; and clinical measurement of body weight. Secondary outcomes were assessed by examining between-arm differences at 6 and 12 months compared with baseline. Further secondary outcomes will be reported elsewhere.

Data Collection Procedures

Data were collected at 4 time points, including screening, baseline, 6 months, and 12 months post randomization. During the screening telephone calls, the research team contacted participants and recorded sociodemographic characteristics (age, sex, education, employment, and language capability), clinical characteristics (duration of NDSS registration and stabilization of the health condition), and app use variables in the REDcap data collection system [20]. Assessments of participants' clinical and psycho-behavioral data were conducted following a standard protocol at baseline and at 6 and 12 months. At each time point, the research team contacted participants in both arms via email to advise that data collection was required. The full data collection details are listed in Multimedia Appendix 2. The email contained a link to the web-based self-administered questionnaire via REDcap. The research team also requested clinical measures from participants and their GPs. If clinical measures were taken within the past 3 months, GPs provided these to the research team; otherwise, participants made an appointment with their GPs to request new tests. GPs provided the clinical information to the research team via fax, and the research team entered the information into REDcap. Data related to program adoption and use were captured automatically by the program (throughout the study period, for all program users), and the information was provided to the research team by Clevertar.

Analytical and Statistical Approaches

On the basis of a previous study [30], we estimated that a sample size of 180 participants would provide 80% power at a two-tailed 5% significance level to detect a difference in HbA_{1c} of 0.4% between the 2 arms with three repeated measures based on a previous study. This estimation assumed a 20% loss to follow-up, a SD change of 1% in HbA_{1c}, and an intraclass correlation coefficient of 0.6 within measurements from the same participant.

All enrolled participants who completed the baseline assessment were included in the *intention-to-treat* analysis [31]. Descriptive analyses were conducted for baseline variables, and *t* tests or chi-square tests were performed (as appropriate) to determine any differences in baseline characteristics between the 2 study arms. For indicators related to program use, mean with SD and range were reported for key metrics. A *chat* was included in the analysis of chat duration only if the whole appointment module with *Laura* was completed during the chat (ie, valid values were available in the program data). Durations were truncated when the records showed extreme values (larger than two interquartile ranges above the third quartile of the distribution), as they were likely to represent cases when participants had forgotten to exit the app and duration had been tracked beyond the end of the actual chat.

The main effectiveness analysis followed the intention-to-treat principle [31]. The analysis fitted a linear mixed-effect model

for repeated measures over the 3 time points of data collection, with a random intercept for the participant and fixed effects for study arm, time point, and study arm-by-time interaction. Two sets of sensitivity analyses were performed: (1) adjusted model: adjusted baseline covariates that were identified as being imbalanced between the two arms or significantly associated with lost-to-follow-up and (2) per-protocol analysis: intervention participants who had completed ≥7 completed chats with *Laura*. Posthoc subgroup analyses were conducted based on age, sex, baseline HbA_{1c}, and NDSS registration duration by adding the subgroup variable and its interaction term with the variable of the interaction between intervention and assessment point as fixed effects to the mixed-effect models used for the main analysis. To further investigate the relationship between program use and effectiveness, mixed-effect models were fitted with a random intercept for the participant and fixed effect for levels of program use (number of chats as 0-6, 7-24, and 25+), assessment time point, and program use-by-time interaction. All analyses were conducted in Stata version 15.1 (Stata Corp) by EG and verified by a statistician (ML) who was blinded to the randomization allocation.

Results

Participant Flow and Baseline Characteristics

Of the 697 individuals with T2DM who expressed interest in participating in the study, 187 were recruited, including 62 (33%) from Victoria, 21 (16.5%) from New South Wales, and 21 (16.5%) from Queensland (Figure 1). The sample included 78 (41.7%) women, with a mean age of 57 (SD 10) years, 76 (40.7%) had a bachelor's degree or above, 88 (47.1%) worked full time, and 72 (38.5%) were registered with the NDSS within the 12 months before recruitment (Table 1). The baseline mean HbA_{1c} was 7.3% (56 mmol/mol, SD 1.5%). Individuals who were excluded were more likely to be females, have relatively lower education levels, be unemployed, and use mobile phones less or do not have access to a smartphone (Multimedia Appendix 2).

The intervention and control arms (n=93 and n=94, respectively) were generally comparable, except that those in the intervention arm were slightly younger (*P*=.04) and had lower depression scores (*P*=.004) at baseline. Among the participants, 116 (62.0%) provided complete data for all 3 assessments (baseline, 6, and 12 months). Incomplete assessments were due to active withdrawal, loss to follow-up, incomplete surveys, or refusal to complete a clinical measure with blood tests. Participants without HbA_{1c} measures provided by GPs at 6 months or 12 months were more likely to have lower baseline HRQoL (*P*=.01) and higher anxiety symptom scores (*P*=.03). Participants who had an incomplete survey without a valid AQoL-8D score were younger (*P*=.02) and more likely to be allocated to the intervention arm (*P*=.007).

Figure 1. Enrollment, randomization, and follow-up of study participants.
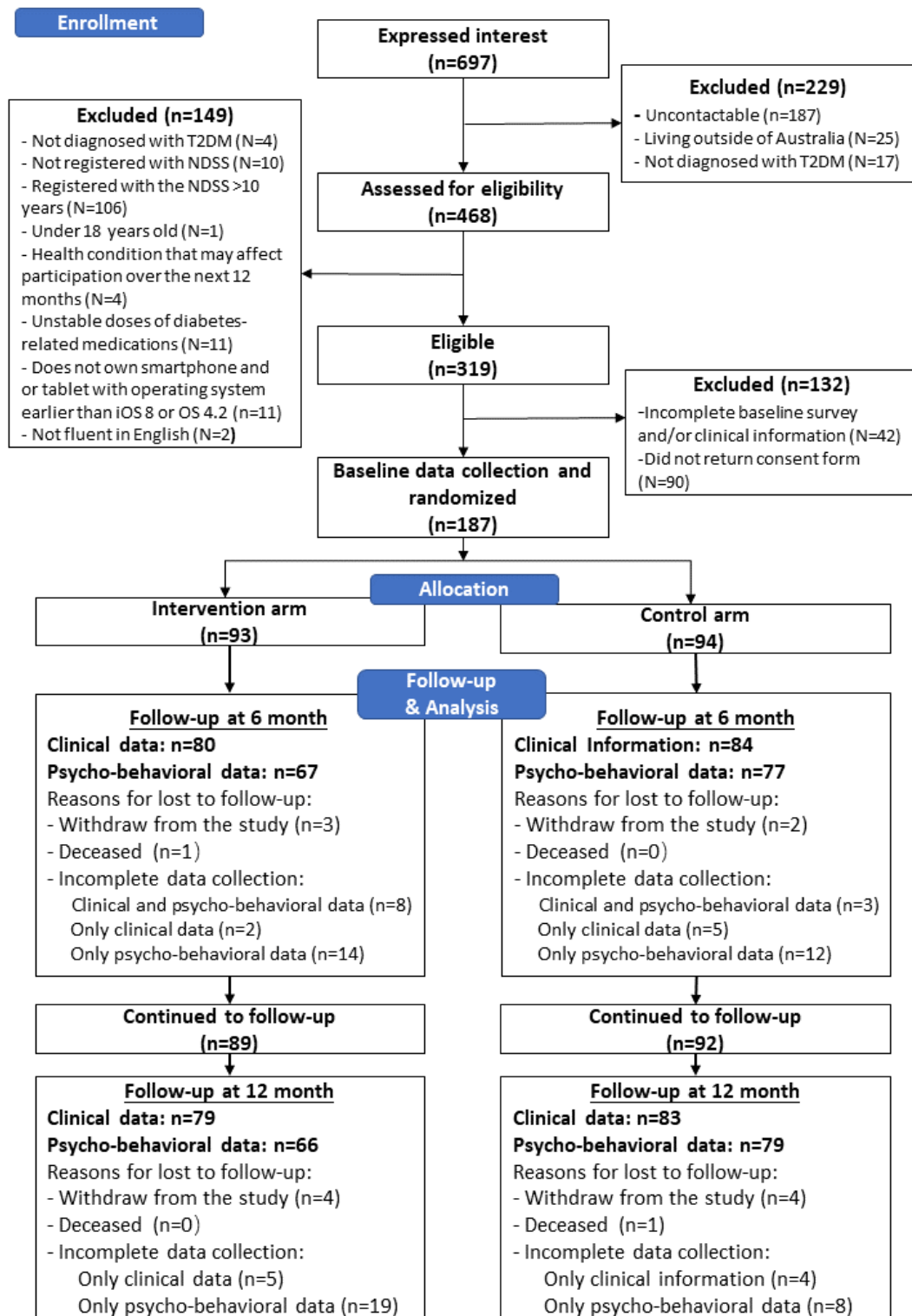
Table 1. Baseline characteristics of study participants

Baseline characteristics	Waitlist (n=94)	Intervention (n=93)	Total (N=187)	P value
Demographic characteristics				
Age[a], (years), mean (SD)	58.4 (10.5)	55.4 (9.7)	56.9 (10.2)	.04
Sex, n (%)				.12
Male	60 (63.8)	49 (52.7)	109 (58.3)	
Female	34 (36.2)	44 (47.3)	78 (41.7)	
Education, n (%)				.25
Secondary high school or lower	29 (30.8)	25 (26.9)	54 (18.8)	
Technical apprenticeship or diploma	30 (31.9)	27 (29.0)	57 (30.5)	
Bachelor's degree	23 (24.5)	17 (18.3)	40 (21.4)	
Postgraduate degree or higher	12 (12.8)	24 (25.8)	36 (19.3)	
Employment, n (%)				.99
Full time	43 (45.7)	45 (48.4)	88 (47.1)	
Part time or casual	16 (17.0)	14 (15.1)	30 (16.0)	
Retired	21 (22.3)	21 (22.6)	42 (22.5)	
Unemployed or others	14 (14.9)	13 (14.0)	27 (14.4)	
English as a secondary language, n (%)	11 (11.7)	6 (6.5)	17 (9.1)	.21
Aboriginal or Torres Strait Islander origin, n (%)	0 (0.0)	4 (4.3)	4 (2.1)	.08
General app use: frequent (multiple times per day), n (%)[b]	67 (71.3)	69 (74.2)	136 (72.7)	.38
Psychosocial characteristics				
Health-related quality of life: Assessment of Quality of Life-8 Dimensions score, mean (SD)	0.7 (0.2)	0.7 (0.2)	0.7 (0.2)	.13
Depressive symptoms: HADS-D[c] score, mean (SD)[a]	4.7 (3.3)	3.3 (3.4)	4.0 (3.4)	.004
Cases (or likely cases) of depression, n (%)	24 (25.5)	12 (12.9)	36 (19.3)	.03
Anxiety symptoms: HADS-A[d] score, mean (SD)	5.6 (3.3)	5.4 (3.8)	5.5 (3.5)	.67
Cases (or likely cases) of anxiety, n (%)	27 (28.7)	28 (30.1)	55 (29.4)	.84
Diabetes-specific distress: PAID[e] score, mean (SD)	30.5 (19.9)	29.2 (21.4)	29.9 (20.6)	.67
Severe diabetes distress (PAID score >40), n (%)	31 (33.0)	29 (31.2)	60 (32.1)	.79
Diabetes-related/health behaviors				
Years registered with NDSS[f], n (%)[b]				.51
≤1 year	34 (38.2)	38 (42.7)	72 (40.5)	
1-5 years	27 (30.3)	23 (25.8)	50 (28.1)	
6-10 years	28 (31.5)	28 (31.5)	56 (31.5)	
Diabetes medications, n (%)				
Diabetes medication(s) prescribed	83 (88.3)	80 (86.0)	163 (87.2)	.64
Insulin prescribed	16 (17.0)	15 (16.1)	31 (16.6)	.06
Taking medicines daily as recommended	71 (75.5)	74 (79.6)	145 (77.5)	.33
Smoking, n (%)	7 (7.5)	4 (4.3)	11 (5.9)	.16
Self-monitoring of blood glucose (>5 days in past 7 days), n (%)[b]	49 (52.1)	53 (57.0)	102 (54.5)	.49
Daily foot checks, n (%)	23 (24.5)	21 (22.6)	44 (23.5)	.87
Clinical characteristics				
Weight (kg), mean (SD)[b]	95.7 (19.0)	97.1 (22.5)	96.4 (20.8)	.65

XSL•FO
RenderX

Baseline characteristics	Waitlist (n=94)	Intervention (n=93)	Total (N=187)	P value
Glycated hemoglobin (%), mean (SD)	7.3(1.6)	7.3(1.5)	7.3 (1.5)	.86
Total cholesterol (mmol/L), mean (SD)[b]	4.5 (1.3)	4.6 (1.4)	4.6 (1.3)	.54
Systolic blood pressure (mm Hg), mean (SD)[b]	130.4 (13.6)	131.1 (14.6)	130.7 (14.1)	.72
Diastolic blood pressure (mm Hg), mean (SD)	78.5 (9.3)	78.4 (9.4)	78.5 (9.3)	.94
Triglyceride (mmol/L), mean (SD)	2.0 (1.3)	1.8 (0.8)	1.9 (1.1)	.26
Diagnosed comorbidities[b]				
High cholesterol, n (%)	59 (62.8)	64 (68.8)	123 (65.8)	.37
Hypertension, n (%)	52 (55.3)	56 (60.2)	108 (57.8)	.45
Arthritis (rheumatoid, osteoarthritis, or other), n (%)[b]	34 (36.2)	22 (23.7)	56 (30.0)	.11
Depression/anxiety/nervous disorder, n (%)	26 (27.7)	26 (28.0)	52 (27.8)	.60
Heart diseases, n (%)	17 (18.1)	17 (18.3)	34 (18.2)	.60
Diabetes-related eye complications, n (%)	12 (12.8)	12 (12.9)	24 (12.8)	.60
Lung diseases, n (%)	11 (11.7)	13 (14.0)	24 (12.8)	.53
Diabetes-related neuropathy, n (%)	11 (11.7)	11 (11.8)	22 (11.8)	.60
Stomach, duodenal, or gastro-intestinal ulcer, n (%)	11 (11.7)	10 (10.8)	21 (11.2)	.59
Cancer, n (%)	6 (6.4)	8 (8.6)	14 (7.5)	.50
Stroke, n (%)	4 (4.3)	10 (10.8)	14 (7.5)	.14
Peripheral vascular diseases, n (%)	8 (8.5)	6 (6.5)	14 (7.5)	.53
Kidney disease, n (%)	5 (5.3)	6 (6.5)	11 (5.9)	.57
Health care service utilization				
Had an appointment with a general practitioner or specialist in the past 12 months, n (%)	93 (98.9)	92 (98.9)	185 (98.9)	.32
Had an appointment with any other health professional (eg, dietician) in the past 12 months, n (%)	54 (57.4)	63 (67.7)	117 (62.6)	.12
Admitted to hospital in the past 6 months, n (%)	21 (22.3)	12 (12.9)	33 (17.6)	.10
Used any other hospital service over the past 6 months that did not involve an admission, n (%)	20 (21.3)	21 (22.6)	41 (21.9)	.80

[a]Significant difference observed between the intervention and control arms.

[b]Some missing values exist: general app use (n=4), years registered with NDSS (n=9), weight (n=8), systolic blood pressure (n=2), diastolic blood pressure (n=2), total cholesterol (n=11), triglyceride (n=13), medication adherence (n=24), self-monitoring of blood glucose (n=6), diagnosed comorbidities (n=1), and health care service utilization (n=1).

[c]HADS-D: Hospital Anxiety and Depression Scale-Depression score.

[d]HADS-A: Hospital Anxiety and Depression Scale-Anxiety score.

[e]PAID: Problem Areas in Diabetes scale.

[f]NDSS: National Diabetes Service Scheme.

Program Adoption and Use

Table 2 shows the program adoption and use among the 93 participants in the intervention arm. During the 12-month program access, 92 participants (98.9%) completed at least one chat with *Laura*, and 83 participants (89.2%) uploaded their blood glucose levels at least once to the app. These 92 participants completed a total of 1942 chats over 12 months of program access and reached a mean of 243 (SD 212) min of chats per person. In general, the program use, including the number of chats and number of blood glucose uploads, reduced over time of the program access (Multimedia Appendix 2). During the 12-month program access, participants who had at least one chat per month with Laura reduced from 87% in the first month (n=81) to 15% (n=14) in the 12th month, and participants who kept recording their blood glucose in the app dropped from 78% (n=73) to 23% (n=21). In addition, the system generated 297 clinical alerts (related to issues on glucose levels, foot ulcer management, and medication side effects) and 179 technical alerts (eg, glucose uploading failed or user did not interact with the app for 14 days) to the program coordinator. A total of 19 discussion topics were posted on the web-based discussion forum covering a broad range of topics such as stigma and social support, depression, and stress management to facilitate the discussion among participants in the intervention arm.

Table 2. Indicators of program adoption and use among participants in the intervention arm (My Diabetes Coach app).

Indicators of program adoption and use	Intervention arm (n=93)
Uptake of the MDC[a] app, n (%)	
Participants who had at least one "appointment" with "Laura" over 12 months	92 (99)
Participants who had uploaded glucose data into the MDC app	83 (89)
Number of completed chats with "Laura" over 12 months, n (%)	
0-6	26 (28)
7-24	37 (40)
25 or more	30 (32)
Total number of chats completed over 12 months	1942
Total number of valid chats completed over 12 months[b]	1641
Individual-level MDC app usage[c], mean (SD); range	
Number of chats completed per person	21.8 (16.7); 1-65
Number of valid chats completed per person[b]	18.4 (15.0); 1-53
Duration of valid chats per person (in minutes), mean (SD); range[d]	
Total duration of chats	242.7 (212.3); 0-1050
Mean duration of each valid chat	13.4 (4.8); 3-26.8
Glucose data uploaded, mean (SD); range	
Number of glucose level uploads per person	181.8 (192.1); 1-966
Program delivery, total; mean per month (SD)	
Clinical alerts (eg, abnormal glucose level)	297; 13.7 (8.8)
Technical alerts (eg, glucose uploading failed)	179; 8.3 (6.5)
Posts on the web-based discussion forum	19; 1.1

[a]MDC: My Diabetes Coach.

[b]Invalid chats were defined as chats for which participants exited the app before the coach modules were fully completed with the closing remark.

[c]For individual-level information, the estimation is based on 92 participants who had records of chat with Laura through the app and 83 participants who had uploaded their glucose levels into the app. Mean (SD) and range of number and duration of chats and glucose data uploads were reported.

[d]Only completed chats have been included in the calculation of the total duration of chats. If the users did not exit the app after completing the chats, the duration would be continuously counted. Thus, we truncated the values if the duration of the chats were more than two interquartile ranges above the third quartile of the distribution.

Program Effectiveness

There was a statistically significant between-arm difference at 12 months in the mean change in HRQoL (AQoL-8D utility value: 0.04, 95% CI 0.00 to 0.07; P=.04) but not in HbA_{1c} (−0.04, 95% CI −0.45 to 0.36; P=.83; Table 3). The MDC arm had a significant improvement in HRQoL from baseline to 12 months (mean estimated change of AQoL-8D score: 0.04, 95% CI 0.01 to 0.06; P=.007). This was maintained from the increase at 6 months (mean estimated change: 0.05, 95% CI 0.01 to 0.08; P=.006). HRQoL remained stable for participants in the control arm between baseline and 12 months (mean estimated change: −0.00, 95% CI −0.03 to 0.02; P=.92). The adjusted model showed consistent results. The per-protocol analysis yielded a slightly larger effect size, as the 12-month between-arm difference in the AQoL-8D score was 0.06 (95% CI 0.02 to 0.09; P=.003). Compared with baseline, the intervention resulted in an increase in the score of the physical health and mental health subscales (Multimedia Appendix 2). There were significant between-arm differences in the mean change of mental health subscore at 6 months as well as the mean change of physical health subscore at 12 months.

Table 3. Effectiveness of the intervention on coprimary outcomes.

Coprimary outcomes and analysis models and Arms	Between arm differences at 6 months (95% CI)	P value	Between arm differences at 12 months (95% CI)	P value	Estimated mean changes between baseline and 6 months (95% CI)[a]	P value	Estimated mean changes between baseline and 12 months (95% CI)[a]	P value
Glycated hemoglobin (%)								
Primary (Intention-to-treat) model[b]								
Intervention	0.06 (−0.35 to 0.47)	.78	−0.04 (−0.45 to 0.36)	.84	−0.20 (−0.49 to 0.09)	.17	−0.33 (−0.62 to −0.04)	.03
Control	Reference	N/A[c]	Reference	N/A	−0.26 (−0.55 to 0.03)	.08	−0.28 (−0.57 to 0.00)	.05
Adjusted model[d]								
Intervention	0.06 (−0.35 to 0.46)	.79	−0.04 (−0.44 to 0.37)	.87	−0.20 (−0.49 to 0.09)	.18	−0.32 (−0.61 to −0.03)	.03
Control	Reference	N/A	Reference	N/A	−0.25 (−0.54 to 0.03)	.09	−0.28 (−0.57 to 0.00)	.05
Per-protocol model[e]								
Intervention	−0.05 (−0.47 to 0.37)	.81	−0.14 (−0.56 to 0.28)	.52	−0.26 (−0.59 to 0.07)	.12	−0.40 (−0.73 to −0.06)	.02
Control	Reference	—	Reference	—	−0.21 (−0.47 to 0.05)	.11	−0.26 (−0.51 to −0.00)	.05
Health Related Quality of Life: Assessment of Quality of Life-8D utility score								
Primary (Intention-to-treat) model[b]								
Intervention	0.05 (0.01 to 0.08)	.006	0.04 (0.00 to 0.07)	.039	0.04 (0.01 to 0.07)	.002	0.04 (0.01 to 0.06)	.007
Control	Reference	N/A	Reference	N/A	−0.01 (−0.03 to 0.02)	.48	0.00 (−0.03 to 0.02）	.92
Adjusted model[d]								
Intervention	0.05 (0.01 to 0.08)	.005	0.03 (0.00 to 0.07)	.047	0.04 (0.01 to 0.06)	.002	0.03 (0.01 to 0.06)	.009
Control	Reference	N/A	Reference	N/A	−0.01 (−0.03 to 0.01)	.46	−0.00 (−0.02 to 0.02)	.93
Per-protocol model[e]								
Intervention	0.06 (0.02 to 0.09)	.002	0.06 (0.02 to 0.09)	.003	0.05 (0.02 to 0.08)	.001	0.05 (0.02 to 0.08)	.001
Control	Reference	N/A	Reference	N/A	−0.01 (−0.03 to 0.01)	.49	−0.01 (−0.03 to 0.02)	.63

[a]Mean changes in outcomes were estimated based on the linear mixed-effect regression model.

[b]For HbA$_{1c}$, the intraclass correlation coefficient (ICC) for the primary model was 0.551 (95% CI 0.465-0.634). For HRQoL, the ICC for the unadjusted model was 0.847 (95% CI 0.806-0.880). Number of participants with valid data at each time point: n for HbA$_{1c}$ (intervention vs control): 93 vs 94 at baseline, 78 vs 78 at 6 months, and 77 vs 79 at 12 months. Number of participants at each time point for HRQoL (intervention vs control): 93 vs 94 at baseline, 67 vs 77 at 6 months, and 60 vs 78 at 12 months.

[c]N/A: not applicable.

[d]The adjusted model adjusted baseline values of variables that were either imbalanced by intervention allocation by chance (baseline age and depression score) or associated with loss to follow-up (baseline AQoL-8D utility value and HADS Anxiety score).

[e]The per-protocol analysis considered participants who had completed more than 6 chats with Laura as following the study protocol.

Compared with baseline, the mean estimated HbA$_{1c}$ decreased in both arms at 12 months (intervention arm: mean estimated change: −0.33%, 95% CI −0.62 to −0.04; P=.03 and control arm: −0.28%, 95% CI −0.57 to 0.00; P=.05). The results were consistent in the adjusted model. In the per-protocol model, a larger between-arm difference was observed at 12 months (between-arm difference: −0.14%, 95% CI −0.56 to 0.28; P=.52).

There was a dose-response relationship between the number of chats and the change in the HRQoL score (Multimedia Appendix

2). Compared with people who completed chats less than 7 times, those who completed more than 24 chats with *Laura* during program access had a significantly greater improvement in AQoL-8D utility score (between-arm difference: 0.09, 95% CI 0.02 to 0.17; P=.02) at 12 months. There was no significant difference in HbA_{1c} (–0.33%, 95% CI –1.11 to 0.44; P=.40).

Although this study was not powered for subgroup analyses, the results did show some statistically significant differences favoring the intervention for HbA_{1c}. There were greater between-arm differences at 6 and 12 months in the mean change in HbA_{1c} for those with higher baseline HbA_{1c} and those registered on the NDSS within the previous year (Multimedia Appendix 2). There were no significant differences in HRQoL at either 6 or 12 months in the sex and age subgroups. The results for secondary outcomes are presented in Multimedia Appendix 2. A significant between-arm difference (–0.89, 95% CI –1.74 to –0.04; P=.04) was observed in the mean change in the HADS anxiety score at 6 months but not at 12 months or for other secondary outcomes reported here.

Discussion

This study is among the very few randomized controlled trials that have evaluated the adoption, use, and effectiveness of a mobile app–based, interactive, embodied conversational agent to support diabetes self-management within home settings over a 12-month period. Our study adds new evidence to this emerging field by demonstrating the program use in a home setting and the effectiveness of app-based interactive conversational agents in supporting diabetes self-management. The MDC program was feasible and shown to be effective in improving participants' HRQoL. Although HbA_{1c} levels reduced during the trial, the between-arm difference was not statistically significant at 6 or 12 months.

Despite the growing number of studies using mobile technologies for diabetes management [10,11,15], the effectiveness of introducing apps to people with T2DM to support their self-management at large scale remains uncertain given the generally poor quality of apps [12,13,32], unmet consumer needs [14], and studies lacking robust designs with long-term evaluations [10,15]. The MDC program is highly innovative with its conversational element. The MDC app incorporated interactive voice recognition and an embodied conversational agent, *Laura*, with human-like characteristics who used a very conversational style of speech to provide people with T2DM with personalized coaching and support on a range of essential diabetes self-management activities in their home environment. The process evaluation that received response from 66 out of 93 participants at 6 months showed that more than 80% of them considered *Laura* as a helpful, friendly and competent assistant and 72% described *Laura* as trustworthy [33].

The indicators for program use suggest successful uptake among individuals with T2DM, but maintaining long-term program use still remains a challenge. Interestingly, program exposure (an average of 18 *appointments* with *Laura*) is very similar to the TLC diabetes program previously developed and evaluated

by our team in Australia [34]; however, the program is more intensive than a human-delivered program that is widely available in Australia (a maximum of eight face-to-face group sessions per calendar year) [35]. Some gamification elements and human-like characteristics were utilized in the MDC program design, as suggested by previous studies [36]. The recently published and separate evaluation from MDC users demonstrated that these techniques did increase users' engagement with the program but were insufficient to ensure engagement among all users [37]. Program users have diverse preferences in terms of program interaction, which is likely determined by their technology literacy, personal preference, and their existing self-management style [37]. The decreasing trend in program use over time and the dose-response relationship between the level of app use and its effectiveness suggest that more efforts are still required to improve the maintenance of program use over time.

We observed a modest but statistically significant improvement in HRQoL at 6 months, which was maintained at 12 months. This finding is consistent with some previous research, which indicates a small but statistically significant benefit from mobile app–based interventions on HRQoL [15]. The improvement in AQoL utility score was relatively small in scale (0.04 increase from baseline to 12 months), but previous literature indicates that even a small effect size for HRQoL may bring substantial well-being benefit in the long term [8,27,38]. In addition, the magnitude of HRQoL improvement is associated with the dose of program use among individuals who accessed the program; this demonstrates the importance of increasing program use in future studies. Such HRQoL improvement was contributed by the improvement in both mental health and physical health, as indicated by the subscale analysis of AQoL. The reduced anxiety symptoms, indicated by the significant between-arm difference in anxiety scores at 6 months, further confirmed the impact of the program on mental health. Although the intervention reduced HbA_{1c}, it did not achieve a significant reduction compared with previous programs [15,39], and there are several likely reasons for this. First, the baseline HbA_{1c} level was not an inclusion criterion, and 52.4% of participants had an HbA_{1c} in the recommended target range (<7% or 53 mmol/mol) at baseline. However, the significant difference by baseline HbA_{1c} observed in the subgroup analysis suggests that the intervention may have significant benefits if the study had targeted only those individuals with HbA_{1c} above target. Second, similar to some previous digital health trials [40], there was a reduction in HbA_{1c} among participants in the control arm. This could be due to the Hawthorne effect with participants sufficiently motivated to be involved in the trial are likely to seek better self-management through other options when they were not allocated to the active intervention. Third, participants' lack of maintenance in program use may have led to a relatively low dose of program exposure. The larger between-arm differences observed in the per-protocol analysis and the greater impact among people with higher program exposure indicates that the effect of the MDC program could have been larger with better fidelity to the study protocol.

This study has several important strengths. First, study participants were recruited from across Australia, with broad

inclusion criteria, and the program was delivered within participants' home settings. Thus, the sample is broadly representative of Australians with T2DM, and the study findings are likely to be generalizable and scalable. Second, the study followed participants over 12 months, which is a relatively long term compared with many studies of this kind [15]. Third, we applied intention-to-treat and per-protocol analysis and performed further analysis of the relationship between program use and effectiveness. As a pragmatic behavioral trial, applying both analysis frameworks provides evidence on the effectiveness of the intervention, both for the ideal situation based on the protocol design and for the real-world situation with adaptation among users.

There were also some limitations. First, due to the nature of the intervention, we were not able to blind participants or their GPs (who provided clinical measurements) to the study arm allocation. Without being able to blind participants, self-report bias and Hawthorne effects may exist. Second, we observed a higher rate of completed assessments among participants in the control arm than in the intervention arm, possibly due to their interest in gaining program access or because of higher attrition in the intervention arm. Third, due to the relatively small sample size, the subgroup analyses should be interpreted with caution. Although the main analyses were fully powered, the subgroup analysis was underpowered and multiple testing would have increased the likelihood of false positives. Overall, the sample was not dissimilar from previous trials in this field [15].

Conclusions

To summarize, this study presents findings concerning the effectiveness, adoption, and use of the MDC program, an app-based interactive embodied conversational agent, in supporting individuals with T2DM. Participants had good adoption of the program and completed a significant amount of chats with *Laura* over 12 months. The trial demonstrated benefits for HRQoL among individuals with T2DM in Australia, which remained apparent at 12 months. The lack of effect for HbA_{1c} is likely due to the relatively low baseline HbA_{1c} level and declining program use over time. Strategies that would increase the program engagement and maintenance in use and improve the effect on HbA_{1c} are required before program scale-up.

Conflicts of Interest

BO and DB received some royalty payments for the development of the scripts for the MDC platform.

Multimedia Appendix 1
Introduction about My Diabetes Coach Program and its five components.
[DOCX File , 2729 KB - jmir_v22i11e20322_app1.docx]

Multimedia Appendix 2
Supplementary tables and figures.
[DOCX File , 229 KB - jmir_v22i11e20322_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 1512 KB - jmir_v22i11e20322_app3.pdf]

References

1. IDF Diabetes Atlaas. International Diabetes Federation. 2019. URL: https://www.diabetesatlas.org/en/resources/ [accessed 2020-06-30]

2. Edelman SV, Polonsky WH. Type 2 diabetes in the real world: the elusive nature of glycemic control. Diabetes Care 2017 Nov;40(11):1425-1432. [doi: 10.2337/dc16-1974] [Medline: 28801473]

3. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]

4. Lindner H, Menzies D, Kelly J, Taylor S, Shearer M. Coaching for behaviour change in chronic disease: a review of the literature and the implications for coaching as a self-management intervention. Aust J Prim Health 2003;9(3):177. [doi: 10.1071/py03044]

5. Pirbaglou M, Katz J, Motamed M, Pludwinski S, Walker K, Ritvo P. Personal health coaching as a type 2 diabetes mellitus self-management strategy: a systematic review and meta-analysis of randomized controlled trials. Am J Health Promot 2018 Sep;32(7):1613-1626. [doi: 10.1177/0890117118758234] [Medline: 29658286]

6. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]

7. Wolever R, Dreusicke M, Fikkan J, Hawkins T, Yeung S, Wakefield J, et al. Integrative health coaching for patients with type 2 diabetes: a randomized clinical trial. Diabetes Educ 2010;36(4):629-639. [doi: 10.1177/0145721710371523] [Medline: 20534872]

8. Williams ED, Bird D, Forbes AW, Russell A, Ash S, Friedman R, et al. Randomised controlled trial of an automated, interactive telephone intervention (TLC Diabetes) to improve type 2 diabetes management: baseline findings and six-month outcomes. BMC Public Health 2012 Aug 3;12:602 [FREE Full text] [doi: 10.1186/1471-2458-12-602] [Medline: 22857017]

9. Gordon L, Bird D, Oldenburg B, Friedman R, Russell A, Scuffham P. A cost-effectiveness analysis of a telephone-linked care intervention for individuals with type 2 diabetes. Diabetes Res Clin Pract 2014 Apr;104(1):103-111. [doi: 10.1016/j.diabres.2013.12.032] [Medline: 24503043]

10. Aminuddin H, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: a systematic review and meta-analysis. Int J Nurs Stud 2019 Feb 8:103286. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]

11. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]

12. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]

13. Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, et al. Quality, functionality, and features of chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 7;8(4):e14836 [FREE Full text] [doi: 10.2196/14836] [Medline: 32255432]

14. Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What do adults with type 2 diabetes want from the 'perfect' app? Results from the second diabetes miles: Australia (MILES-2) study. Diabetes Technol Ther 2019 Jul;21(7):393-399. [doi: 10.1089/dia.2019.0086] [Medline: 31166804]

15. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]

16. Garabedian LF, Ross-Degnan D, LeCates RF, Wharam JF. Uptake and use of a diabetes management program with a mobile glucometer. Prim Care Diabetes 2019 Dec;13(6):549-555. [doi: 10.1016/j.pcd.2019.03.010] [Medline: 31097343]

17. Shen Y, Wang F, Zhang X, Zhu X, Sun Q, Fisher E, et al. Effectiveness of internet-based interventions on glycemic control in patients with type 2 diabetes: meta-analysis of randomized controlled trials. J Med Internet Res 2018 May 7;20(5):e172 [FREE Full text] [doi: 10.2196/jmir.9133] [Medline: 29735475]

18. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]

19. Australian Diabetes Map. National Diabetes Service Scheme. 2019. URL: https://www.ndss.com.au/about-the-ndss/diabetes-facts-and-figures/australian-diabetes-map/ [accessed 2020-06-30]

20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

21. Prochaska J, Redding C, Evers K. The transtheoretical model and stages of change. In: Health Behavior: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2015:125-148.

XSL•FO

RenderX

22. Crosby R, Salazar L, DiCelemente R. Social cognitive theory applied to health behavior. In: Health Behavior Theory for Public Health: Principles, Foundations, and Applications. Burlington, UK: Jones & Bartlett Learning; 2013:163-186.

23. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]

24. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]

25. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]

26. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the assessment of quality of life (AQoL)-8D multi-attribute utility instrument. Patient 2014;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]

27. Hawthorne G, Osborne R. Population norms and meaningful differences for the assessment of quality of life (AQoL) measure. Aust N Z J Public Health 2005 Apr;29(2):136-142. [doi: 10.1111/j.1467-842x.2005.tb00063.x] [Medline: 15915617]

28. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]

29. Welch GW, Jacobson AM, Polonsky WH. The problem areas in diabetes scale. An evaluation of its clinical utility. Diabetes Care 1997 May;20(5):760-766. [doi: 10.2337/diacare.20.5.760] [Medline: 9135939]

30. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]

31. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. Int J Epidemiol 1992 Oct;21(5):837-841. [doi: 10.1093/ije/21.5.837] [Medline: 1468842]

32. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]

33. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. Acceptability of an embodied conversational agent for type 2 diabetes self-management education and support via a smartphone app: mixed methods study. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17038 [FREE Full text] [doi: 10.2196/17038] [Medline: 32706734]

34. Bird D, Oldenburg B, Cassimatis M, Russell A, Ash S, Courtney MD, et al. Randomised controlled trial of an automated, interactive telephone intervention to improve type 2 diabetes self-management (telephone-linked care diabetes project): study protocol. BMC Public Health 2010 Oct 12;10:599 [FREE Full text] [doi: 10.1186/1471-2458-10-599] [Medline: 20937148]

35. Group Allied Health Services Under Medicare For Patients With Type 2 Diabetes Patient Information. Australian Government Department of Health. 2013. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/D89B8F9BD812E58CCA257BF000209C70/$File/Group%20Allied%20Health%20Services-Type%202%20Diabetes-Patient%20Info.pdf [accessed 2020-06-30]

36. Priesterroth L, Grammes J, Holtz K, Reinwarth A, Kubiak T. Gamification and behavior change techniques in diabetes self-management apps. J Diabetes Sci Technol 2019 Sep;13(5):954-958 [FREE Full text] [doi: 10.1177/1932296818822998] [Medline: 30762420]

37. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]

38. Cochran J, Conn VS. Meta-analysis of quality of life outcomes following diabetes self-management training. Diabetes Educ 2008;34(5):815-823 [FREE Full text] [doi: 10.1177/0145721708323640] [Medline: 18832286]

39. Captieux M, Pearce G, Parke HL, Epiphaniou E, Wild S, Taylor SJ, et al. Supported self-management for people with type 2 diabetes: a meta-review of quantitative systematic reviews. BMJ Open 2018 Dec 14;8(12):e024262. [doi: 10.1136/bmjopen-2018-024262] [Medline: 30552277]

40. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]

Abbreviations

AQoL-8D: Assessment of Quality of Life-8 Dimensions
CONSORT: Consolidated Standards of Reporting Trials
GP: general practitioner
HADS: Hospital Anxiety and Depression Scale
HbA$_{1c}$: glycated hemoglobin
HRQoL: health-related quality of life
MDC: My Diabetes Coach
NDSS: National Diabetes Service Scheme

NHMRC: National Health and Medical Research Council
T2DM: type 2 diabetes mellitus
TLC diabetes: Telephone-Linked Care for Diabetes

XSL•FO
RenderX

Original Paper

Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study

Jason Koman[1,2,3*], PhD; Khristina Fauvelle[4], MD; Stéphane Schuck[1*], MD; Nathalie Texier[1], MD; Adel Mebarki[1], MD

[1]Kap Code, Paris, France

[2]CNRS, PASSAGES, Bordeaux, France

[3]Bordeaux Population Health Research Center, University of Bordeaux, Inserm, Bordeaux, France

[4]Sanofi Aventis, Gentilly Cedex, France

[*]these authors contributed equally

Corresponding Author:
Khristina Fauvelle, MD
Sanofi Aventis
82, avenue Raspail
Gentilly Cedex, 94255
France
Phone: 33 772219558
Email: khristina.fauvelle@sanofi.com

Abstract

Background: Seeking medical information can be an issue for physicians. In the specific context of medical practice, chatbots are hypothesized to present additional value for providing information quickly, particularly as far as drug risk minimization measures are concerned.

Objective: This qualitative study aimed to elicit physicians' perceptions of a pilot version of a chatbot used in the context of drug information and risk minimization measures.

Methods: General practitioners and specialists were recruited across France to participate in individual semistructured interviews. Interviews were recorded, transcribed, and analyzed using a horizontal thematic analysis approach.

Results: Eight general practitioners and 2 specialists participated. The tone and ergonomics of the pilot version were appreciated by physicians. However, all participants emphasized the importance of getting exhaustive, trustworthy answers when interacting with a chatbot.

Conclusions: The chatbot was perceived as a useful and innovative tool that could easily be integrated into routine medical practice and could help health professionals when seeking information on drug and risk minimization measures.

(J Med Internet Res 2020;22(11):e15185) doi:10.2196/15185

KEYWORDS

health; digital health; innovation; conversational agent; decision support system; qualitative research; chatbot; bot; medical drugs; prescription; risk minimization measures

Introduction

A conversational agent, also known as a "chatbot," is an artificial system that can converse with a human user through automated message exchange [1]. In order to interact in a natural way with the user, it employs a question and answer database [2]. Commonly used in marketing as a consumers' guide, chatbots also have been developed in the health field in several applications aimed mostly towards patients [3-11]. However, little scientific research has examined the use of conversational agents through physicians' viewpoints [12-15], and none have taken into account medical information addressed to health professionals. It has been shown that nearly all physicians use the internet for medical information seeking [16,17] but have to cope with an increasing flow of medical information [18]. Too much information to scan could be a barrier to respond to a defined medical question [16,17]; thus, seeking health care information can be an issue for physicians since their information sources can be fragmented, incomplete, and not easy to find [19]. Credibility of the source, relevance, unlimited

XSL•FO
RenderX

access, speed, and ease of use are the main criteria for information seeking among physicians [17]. A national survey conducted in France in 2017 showed that the prescription and delivery methods of drugs could be incorrectly followed by health professionals, despite the risk minimization measures campaign conducted at a national level [20]. Providing medical information to health professionals through a conversational agent could help to detect pharmacovigilance cases early and to reinforce the proper intake of medication by patients. More specifically, chatbots could be deployed as a complementary solution to provide information about drugs with the purpose of minimizing drug risks. In France, drug risk minimization measures are generally recommended by the European Pharmacovigilance Risk Assessment Committee and adapted by the French National Agency for Medicines and Health Products Safety (Agence Nationale de Sécurité du Médicament et des Produits de Santé) to regulate medical practice and ensure safe, efficient intake of drugs [21].

To improve drug information access and awareness of drug risk minimization measures among physicians, a pilot version of a chatbot was developed and tested. The chatbot was expected to improve drug information access and awareness of drug risk minimization measures. This qualitative study aimed to improve the chatbot to meet physicians' needs and expectations. User participation [22] was used to elicit physicians' perceptions of the chatbot through the use of the pilot version. Gathering more in-depth qualitative information on these topics may be useful to help develop a conversational agent that meets physicians' needs and expectations.

Methods

Study Design

The study focused on the pilot version of a text-based chatbot developed by an international pharmaceutical company. This was qualitative research that employed individual, in-depth, semistructured interviews to explore physicians' opinions and perceptions of a chatbot. The aim of the chatbot was to provide quick, 24-hour access to information on drugs for physicians to improve their medical practice and the application of recommendations regarding drug risk minimization measures among patients. Thus, all information provided by the chatbot was sourced from official regulatory documents issued by French health authorities to ensure patient security. The chatbot was accessible online through a web browser. When logged in, it displayed a dialog box on the right side of the screen and a female figure in the central area. Answers provided by the chatbot followed pre-established topics related to drugs (eg, treatment initiation, discontinuing treatment, treatment renewal, documentation or information that can be downloaded and obtained by the user). When asked a question, the chatbot provided a direct answer based on its database. When a question was not sufficiently focused, the chatbot proposed between 2 and 4 categories related to the question. The user then had to choose between those categories by typing the corresponding number. When the chatbot was asked a question that was not available in its database or not understood because of a typing error, the user was alerted. In this latter case, the user had to

reformulate the question or ask a new question related to pre-established topics. As a pilot version, the chatbot initially focused on one drug. Further information can be added and delivered when the final version is deployed; the chatbot is meant to be used by physicians who seek quick access to information in the presence of a patient or when preparing a consultation.

General practitioners were approached initially as the conversational agent was developed for this specific population. Specialists were also approached because some of their patients used drugs subject to risk minimization measures.

Recruitment and Participants

Participants were recruited on a voluntary basis after being asked by email whether they were interested in "taking part in a project that will test an innovative digital health tool." Their email addresses were available from a database of health practitioners who had already participated in previous digital epidemiological studies. The interviewer had not met the participants before the interview. General practitioners were recruited to represent various situations such as urban and rural general practice. Specialists were recruited from either public or private practice. Thematic saturation was not taken into account since the study aimed to improve the chatbot through physicians' needs and expectations in an explorative way without being representative.

Data Collection

Interview Design

Semistructured interviews were conducted during June 2018 and July 2018 and took place primarily in participants' medical offices. At the beginning of the interview, each participant received a brief summary of the study from the interviewer, and sociodemographic data (gender, age, profession, and workplace) were collected. Interviews were scheduled for approximatively 1 hour.

To standardize the interviews, they were divided into 3 sections: introduction, introduction of the chatbot, and conclusion.

Introduction

The introduction included knowledge about the use of drugs and recommendations about risk minimization measures and acceptability of a chatbot within their daily practice.

Presentation of the Chatbot

The chatbot was presented through the example of 3 preselected input phrases based on various situations: "How to accompany a pregnant woman taking this medicine?", "I want to renew the prescription of a patient.", "Does this drug decrease the impact of contraception?" Test sessions were conducted as follows. After the introduction, the interviewer logged in to the chatbot with a confidential password. In the dialog box, introductory text was displayed and read aloud by the interviewer to explain the purpose of the chatbot. Then, preselected input phrases were first inputted by the interviewer to show how the chatbot worked. Participants were then asked to try the chatbot by searching for information on drugs as they would do during a real consultation, using their own terminology. Questions were

collected by the interviewer after a physician concluded the test, but participants were left free to use the conversational agent as much as they wanted during the rest of the interview. Since it was a pilot version, participants did not have access to the conversational agent before or after the interview. Moreover, the test was different from a real-case scenario since participants were not totally autonomous in using the chatbot and had to share their experience with the interviewer.

Conclusion

During the conclusion, feedback was collected on the user experience with the chatbot. Physicians were asked about the relevance and appreciation of the tested tool as well as their needs for and expectations of the chatbot.

Template

The interviewer followed a semistructured interview template composed of the following themes: (1) knowledge about drug risk minimization measures and acceptability of digital health within their daily practice (Section 1), (2) relevance and appreciation of the chatbot to facilitate information acquisition (Section 3), and (3) needs for and expectations of the tested chatbot (Section 3).

Data Analysis

Interviews were recorded using a digital audio recorder after receiving participant agreement. They were fully transcribed and anonymized by the interviewer. Transcripts were organized, sorted, and coded using a systematic thematic analysis approach [23]. Main themes were developed and identified by the authors as patterns emerged within data [24]. Then, they were graphically represented through the use of Visual Understanding Environment (VUE) software. For each participant, data were summarized in one table following the main themes identified during the interviews. Quotations that were not related to the chatbot were excluded from analysis. Data from each theme were gathered in another table to proceed with the horizontal thematic analysis by highlighting sentences expressed by participants on each topic.

Results

Demographics

A total of 10 health professionals participated in individual, in-depth, semistructured interviews in France.

The sample size included 8 general practitioners and 2 specialists. The average age of the participants was 51.5 years. There were 7 men and 3 women; 6 participants worked in an urban area only, 3 worked in a rural area only, and 1 worked in both urban and rural areas. All participants were following patients who needed routine drug prescriptions including risk minimization measures.

All participants reported being vigilant about new recommendations for prescription including risk minimization measures before using the chatbot. Seven physicians reported that they sought information on drugs subject to risk minimization measures only when their patients were directly

concerned, when they were confronted with a particular case, or when adverse events had been reported to them:

> *I have access to information obviously, but I get more interested when I am confronted with the problem.* [Man, in practice for 35 years]

They considered the information on drug risk minimization measures easy to find but fragmented. The main source of information they reported was the internet, followed by information from health authorities (letters, email), databases on drugs, medical journals, and communications from laboratories (press, visits). Exchanges with other practitioners had also been cited as sources of information:

> *The easiest and most accessible information is the summary of characteristic products, but it is in rather indigestible to read since it's really big.* [Man, in practice for 11 years]

> *Generally, I do research spontaneously. I'm looking out on the internet when I have a question about a treatment or when I'm not sure about a question a patient asked me, we can look together.* [Woman, in practice for 8 years]

> *I get information from National Agency for Medicines and Health Products Safety, from health authorities, and from laboratories as well. I receive either emails or letters.* [Man, in practice for 30 years]

Central Themes

Emerged Topics

Several themes and subthemes emerged from the interviews. The central themes were (1) achievement by man including ergonomics and format of asking; (2) achievement by tool including design, tone, and form of presentation; (3) content of output including the amount of information, clarity, and accuracy; and (4) user needs. These topics are discussed with illustrative quotes in the following paragraphs.

Achievement by Man

In total, 52 questions were asked to the chatbot by participants. Of the 52 questions formulated by the physicians (Multimedia Appendix 1) during tests, 24 were answered by the pilot version of the conversational agent. One-third of the 24 answers (7/24) were obtained on the first try, while the other two-thirds (17/24) needed a complementary question. Furthermore, 28 questions were not understood by the conversational agent, and 8 questions were not included in the conversational agent themes. Questions addressed to the conversational agent were about prescription, medical treatment renewal, drug side effects, drug interactions, and records on drug products.

Some physicians judged the ergonomics favorably due to simplicity and ease of use:

> *This is correct. The style is sober, quiet; it is not aggressive. It is simple and readable.* [Man, in practice for 36 years]

> *Yes, it is really easy to use. The ergonomics are really good.* [Man, in practice for 11 years]

Over half (6/10) of the participants found that the chatbot lacked intuitiveness and fluidity to access information:

> *The answer is clear, but I don't know, there is something with the fluidity... If I compare with where I'm usually looking for information, there is one section for each topic, and I get instant access. Here, the answers are too robotic.* [Woman, in practice for 24 years]

> *I would say this is not intuitive. Everyone knows how to use a computer. But the ergonomics... I don't know, there is just a login page, then the home screen. If it is just that, there is no ergonomics. It would be nice to have a website plan.* [Male, in practice for 30 years]

The size of the dialog box was also cited as improvable:

> *There is a part of the screen that is not used; the dialog box looks really small compared to the rest. It could occupy a larger space, which would allow a bigger typeface.* [Man, in practice for 11 years]

> *The window to type our questions is too small. If you enter a long sentence, you can't see the beginning of your sentence, which is bothering. Indeed, the window is too small and the typeface as well.* [Man, in practice for 36 years]

Achievement by Tool

Most of the participants reported that the colors and icon of the chatbot were too simple, but the global design was judged serious and professional, which was appreciated:

> *You could give us the possibility to customize the design by choosing between 3 or 4 different colors. Maybe it should be a bit more cheerful. It looks seriously made, but a little too dark.* [Man, 18 years of practice]

> *I like the font; it is quite nice. It looks as if everything was made to be relatively neutral. This is good because the tool can't be visually aggressive during a consultation. This is not really attractive; it is quite neutral, but this is adapted because visual elements catching your eyes all the time can be tiring. This is better.* [Man, in practice for 11 years]

Seven participants reported that the tone of the answers was positive. However, the display of the answers provided by the conversational agent needed to be more concise according to half (5/10) of the physicians:

> *It speaks from a medical point of view so there is no problem with the tone.* [Man, in practice for 36 years]

> *It needs a display with bold characters so we can reach key points by reading through it diagonally* [Man, in practice for 23 years]

Content of Output

The response delay from the chatbot was acceptable for 9 of 10 participants, but 5 physicians were not satisfied with the length of the information, which was too broad to read or to access:

> *It is a little bit long because every time it asked me for a clarification, it seems that it doesn't understand.* [Woman, specialist in practice for 24 years]

> *This is accurate but it is too much talking. Two lines should be enough; this is too long in my opinion.* [Woman, in practice for 24 years]

Even though all participants found the information provided by the chatbot very clear, the main issue pointed out by physicians was the accuracy of the answers. Obtained answers did not provide a sufficient amount of information. Thus, even if they trusted the provided information, it was considered as too generic and not relevant enough:

> *We need clear and concise scientific information. We need to know globally what to do; then, we translate the information but at least it needs to be precise. Because for now, it is a little bit general; this is too elementary.* [Man, in practice for 35 years]

> *The answers are understandable; it comes out quickly; it comes out clear. The only inconvenience is that it is not the answer I was looking for.* [Man, in practice for 18 years]

> *It is centered on the developers' insight, not from a physician's insight. This is almost a tool for a patient actually. Physicians have much more questions than the chatbot can answer for now.* [Man, in practice for 11 years]

User Needs

Participants identified pros and cons of the chatbot after they tested it (Table 1). Pros were the concept, ergonomics, assistance for diagnosis, and time savings. Cons were the natural language comprehension issues, data security, and the threat to health care professions.

Table 1. Pros and cons of the chatbot.

Observation and examples	Example excerpts
Pros	
Relevant concept: give fast and precise information	"In my opinion, the concept is really interesting. Why? Because this is provisioning a tool in which we can get precise and fast information. This is a concept that seems very interesting and very relevant to me." [Man, in practice for 35 years]
Ergonomic tool: can be utilized as a research tool	"The ergonomics, the conciseness…Yes, the shape. You call this a conversational agent; to me, it is a good research engine." [Man, in practice for 11 years]
Diagnosis helping: can provide assistance to medical care	"The fact that it can give answers when we have a question is interesting. That doesn't mean we'll necessarily follow the answer but at least it gives a direction. Tools that can help medical care or to diagnose, I find great." [Woman, in practice for 24 years]
Time-saving	"There are only benefits anyway. This is instantaneous, verified information. It saves us time. A consultation is time-consuming; we need to get immediate answers. It has already happened to me that I had to call a laboratory to get an answer, and I was told that someone would call me back. Here, I get my answer almost directly" [Man, in practice for 30 years]
Cons	
Natural language comprehension issues: can be difficult to access the right information	"The tree structure to get to the information. That is to say… If I get some information, in what degree is it intuitive? Does it need to be questioned in a precise way, or can I ask my questions without thinking about the formulation? Can it answer my everyday needs?" [Man, in practice for 18 years]
Medical data issue: can be dangerous for medical data security	"Of course, there is hacking. I can't take control of a doctor, but I can get control of a conversational agent without problems. All the security part, as powerful it can be, can always be breached if someone intends to do it. That's the limit." [Man, in practice for 23 years]
Threatens health care professionals: a chatbot could replace doctors	"Let's say that, in 10 years, we are told that our job will be replaced; it is a nightmare. We studied for 10 years, not 2 years, but 10 years. If we are told that we will be replaced by artificial intelligence, it does not make anyone dream." [Man, in practice for 23 years]

Nine participants estimated that this tool would be easily integrated in their routine practice when fully developed, especially if it could provide fast, trustful, easily accessible information on a range of medical topics:

> It is simple; it is practical. When it has acquired more vocabulary, I think it will be an efficient tool. [Woman, in practice for 8 years]

> This is a tool in which we can get precise and fast information. This is a concept that seems very interesting and very relevant to me" [Man, in practice for 35 years]

However, half (5/10) of the participants considered that the chatbot did not meet their needs. The main reason was the natural language comprehension issues:

> It needs to be accurate and well-edocumented because, during a consultation and in our daily routine, we don't have much time. Not having enough time means we need to get to the point; so, we need

> fast, specific information" [Man, in practice for 35 years]

> If I get information, what is the degree of intuitiveness? The machine has to adapt, because when I'm doing my job with a patient in front of me, I can't be focused on how I should formulate my question to the chatbot. [Man, in practice for 18 years]

> During a consultation, everything has to be done quickly, I need information quickly. If the chatbot doesn't understand me, I put it aside to get the information elsewhere. [Woman, in practice for 24 years]

Participants reported areas to improve the chatbot prototype (Textbox 1). Suggestions were made about the content of the conversational agent, its display and appearance, and the extensions of use that could be made. Suggestions included highlighting the most important information provided by the chatbot, the possibility to provide an information sheet to the patients, or integrating medical themes beyond drug risk minimization measures.

Textbox 1. Suggestions made by general practitioners and specialists for improvement of the conversational agent.

Display and appearance

- Highlight important information in the answers

- Possibility to choose interface colors

Content

- Provide [printed] information to the patients

- Regular updates

Extensions of use

- Ask questions orally

- Integrate more themes (eg, drug interactions, pathologies)

- Declare side effects

Discussion

Principal Findings

The study was conducted to elicit physicians' perceptions of a chatbot that is meant to improve drug information access and awareness of drug risk minimization measures. A qualitative approach was chosen to collect detailed data on how the chatbot was used and perceived by physicians on a small scale and needs and expectations for the chatbot. As a pilot version, the chatbot did not meet physicians' expectations.

Overall, the findings demonstrated that physicians' needs towards information delivered by the chatbot were the reliability of sources, precise information, and speed of access. This is consistent with previous research demonstrating that all these features are critical to physicians when seeking information [17]. Physicians were particularly vigilant to the conversational agent content. Even though they did appreciate getting information on risk minimization measures and drugs, they had difficulties accessing the right information because of natural language understanding issues. As a tool based on artificial intelligence, physicians also expected that the chatbot would understand natural language. Previous research demonstrated that over half of physicians believed that chatbots lack the intelligence or knowledge to accurately assess the user [13], which can be an issue regarding speed and ease of use when using the chatbot.

The chatbot format was also appreciated, as it could easily be integrated into physicians' routine practice, either during or before a consultation with a patient. This was judged as innovative since most chatbots in health care are developed to provide information for patients [3-11]. Access to information is also considered easier with a conversational agent than with a classic drug reference database. First, it was considered to provide reliable information in a practical way. Second, it made information accessible so that physicians could quickly find what they were looking for and, consequently, save time. In that way, using a chatbot can improve medical care from physicians' perspectives. Shared decision making may increase the effectiveness of a treatment if the patient is given a sufficient and appropriate amount of information [25]. In this regard, the chatbot could help medical care.

It was pointed out that the conversational agent did not always respond to the questions they formulated, either because it did not recognize the everyday language employed by the users or because the answers, while based on official regulatory documents, were too broad without highlighting the most important parts of information. However, they were aware that the chatbot was still in an experimental stage. Regarding this fact, most practitioners were willing to use the conversational agent in its fully developed version if this tool remains easy to use, secure, and easily accessible.

Areas of improvement for the conversational agent proposed by medical practitioners included a better understanding of the questions formulated, highlighting the most important information, and better ergonomics.

Limitations

This market research is not without limitations. As an exploratory study with a sample size of 10 medical practitioners, findings are not generalizable to the entire population of medical practitioners. This was a qualitative study conducted with semistructured interviews, which allowed us to explore tendencies and opinions on the usage of a conversational agent. The average age of our sample (51.5 years) was slightly above the average age of the medical practitioner population in France, which was 50.8 years in 2018 [19]. The majority of our sample was male practitioners (7/10, 70%), which does not reflect the distribution between male (53%) and female (47%) medical practitioners in France in 2018 [19].

In addition, medical practitioners who were recruited had already participated in other digital epidemiological studies carried out by the same research team, and it is possible that they were more receptive to new health technology. In other words, while our sampling method was not meant to be representative, this specific market research may be prone to researcher bias.

Finally, because of the confidentiality policy, the drug associated with the chatbot could not be cited in the paper. However, we believe that it did not interfere with the results, which were meant to elicit physicians' perceptions of the chatbot.

Conclusions

According to the results of this study, it appears that chatbots could be a solution for quick and easily accessible information. By reinforcing the knowledge on drugs based on official and institutional recommendations, a chatbot could be used to enhance compliance with the drug risk minimization measure within the physician population. In particular, this chatbot prototype was perceived by medical practitioners as a useful, acceptable, innovative tool that could easily be integrated in their daily medical practice. Finally, even though the chatbot prototype could not be used as it was because of insufficient information in the database, findings suggest that physicians are willing to use a chatbot not only for prescription but also to get information on drug interactions or to obtain assistance within medical care for complex pathology or disease management. A future challenge for the chatbot should be to accommodate physicians' needs for accurate, concise information based on official regulatory documents that ensure patients' security.

Multimedia Appendix 1
Questions tested by physicians.
[DOCX File , 19 KB - jmir_v22i11e15185_app1.docx]

References

1. Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A, Sansone C. Chatbots Meet eHealth: Automatizing Healthcare. 2017 Presented at: Workshop on Artificial Intelligence with Application in Health; November 14, 2017; Bari, Italy URL: http://ceur-ws.org/Vol-1982/paper6.pdf

2. Corti K, Gillespie A. A truly human interface: interacting face-to-face with someone whose words are determined by a computer program. Front Psychol 2015;6:634 [FREE Full text] [doi: 10.3389/fpsyg.2015.00634] [Medline: 26042066]

3. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]

4. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health. Front Psychol 2017;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

5. Lokman A, Zain J, Komputer F, Perisian K. Designing a Chatbot for Diabetic Patients. 2009 Presented at: International Conference on Software Engineering & Computer Systems; October 2009; Kuantan, Malaysia p. 19-21.

6. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]

7. Comendador BEV, Francisco BMB, Medenilla JS, Nacion SMT, Serac TBE. Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. JOACE 2015;3(2):137-140. [doi: 10.12720/joace.3.2.137-140]

8. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations. Stud Health Technol Inform 2016;227:55-60. [Medline: 27440289]

9. Stieger M, Nißen M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: study protocol of a randomized, wait-list controlled trial. BMC Psychol 2018 Sep 04;6(1):43 [FREE Full text] [doi: 10.1186/s40359-018-0257-9] [Medline: 30180880]

10. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]

11. Hess GI, Fricker G, Denecke K. Improving and Evaluating eMMA's Communication Skills: A Chatbot for Managing Medication. Stud Health Technol Inform 2019;259:101-104. [Medline: 30923283]

12. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Best practices for designing chatbots in mental healthcare? A case study on iHelpr. Swindon; 2018 Presented at: 32nd International BCS Human Computer Interaction Conference (HCI); July 2-6, 2018; Belfast, Northern Ireland. [doi: 10.14236/ewic/hci2018.129]

13. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]

14. Feldman MJ, Hoffer EP, Barnett GO, Kim RJ, Famiglietti KT, Chueh HC. Impact of a computer-based diagnostic decision support tool on the differential diagnoses of medicine residents. J Grad Med Educ 2012 Jun;4(2):227-231 [FREE Full text] [doi: 10.4300/JGME-D-11-00180.1] [Medline: 23730446]

15. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]

16. Casebeer L, Bennett N, Kristofco R, Carillo A, Centor R. Physician Internet medical information seeking and on-line continuing education use patterns. J Contin Educ Health Prof 2002;22(1):33-42. [doi: 10.1002/chp.1340220105] [Medline: 12004639]

17. Bennett NL, Casebeer LL, Kristofco RE, Strasser SM. Physicians' Internet information-seeking behaviors. J Contin Educ Health Prof 2004;24(1):31-38. [doi: 10.1002/chp.1340240106] [Medline: 15069910]

18. Nylenna M, Aasland O. Primary care physicians and their information-seeking behaviour. Scand J Prim Health Care 2000 Mar;18(1):9-13 [FREE Full text] [doi: 10.1080/02813430050202488] [Medline: 10811036]

19. Begaud B, Costagiola D. Rapport sur la surveillance et la promotion du bon usage du medicament en France. French Ministry of Solidarity and Health. 2013. URL: https://solidarites-sante.gouv.fr/IMG/pdf/Rapport_Begaud_Costagliola.pdf [accessed 2019-02-21] [WebCite Cache ID 76MHfSLD1]

20. Sanofi Aventis France. Evaluation auprès des pharmaciens d'officine des mesures de minimisation du risqué consistant en des nouvelles conditions de prescription et de délivrance (CPD) du valproate. Synthesis from National Agency for Medicines and Health Products Safety. 2017. URL: https://www.ansm.sante.fr/var/ansm_site/storage/original/application/cfc9902b616321c0fc3bc742dc57053b.pdf [accessed 2019-03-12] [WebCite Cache ID 76oseOTZv]

21. Mesures additionnelles de minimisation des risques. National Agency for Medicines and Health Products Safety. 2017. URL: https://www.ansm.sante.fr/Activites/Surveillance-des-medicaments/Mesures-additionnelles-de-reduction-du-risque/(offset)/1 [accessed 2019-02-18] [WebCite Cache ID 76HXB1P43]

22. Barki H, Hartwick J. Measuring User Participation, User Involvement, and User Attitude. MIS Quarterly 1994 Mar;18(1):59-82 [FREE Full text] [doi: 10.2307/249610]

23. Guest G, MacQueen KM, Namey EE. Introduction. In: Applied Thematic Analysis. Thousand Oaks, CA: Sage Publications, Inc; Nov 09, 2011:3-20.

24. Richards L. Readme First For A User's Guide To Qualitative Methods. Thousand Oaks, CA: Sage Publications, Inc; 2019.

25. Coulter A, Entwistle V, Gilbert D. Sharing decisions with patients: is the information good enough? BMJ 1999 Jan 30;318(7179):318-322 [FREE Full text] [doi: 10.1136/bmj.318.7179.318] [Medline: 9924064]

Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study

Keith E Morse[1], MD, MBA; Nicolai P Ostberg[2], MS; Veena G Jones[3,4], MD; Albert S Chan[3,4], MD, MS

[1]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, United States
[2]Center for Biomedical Informatics Research, Stanford University School of Medicine, Palo Alto, CA, United States
[3]Clinical Leadership Team, Sutter Health, Sacramento, CA, United States
[4]Palo Alto Medical Foundation Research Institute, Palo Alto, CA, United States

Corresponding Author:
Keith E Morse, MD, MBA
Department of Pediatrics
Stanford University School of Medicine
750 Welch Road
Suite 315
Palo Alto, CA, 94304
United States
Phone: 1 650 723 5711
Email: kmorse@stanfordchildrens.org

Abstract

Background: Pressure on the US health care system has been increasing due to a combination of aging populations, rising health care expenditures, and most recently, the COVID-19 pandemic. Responses to this pressure are hindered in part by reliance on a limited supply of highly trained health care professionals, creating a need for scalable technological solutions. Digital symptom checkers are artificial intelligence–supported software tools that use a conversational "chatbot" format to support rapid diagnosis and consistent triage. The COVID-19 pandemic has brought new attention to these tools due to the need to avoid face-to-face contact and preserve urgent care capacity. However, evidence-based deployment of these chatbots requires an understanding of user demographics and associated triage recommendations generated by a large general population.

Objective: In this study, we evaluate the user demographics and levels of triage acuity provided by a symptom checker chatbot deployed in partnership with a large integrated health system in the United States.

Methods: This population-based descriptive study included all web-based symptom assessments completed on the website and patient portal of the Sutter Health system (24 hospitals in Northern California) from April 24, 2019, to February 1, 2020. User demographics were compared to relevant US Census population data.

Results: A total of 26,646 symptom assessments were completed during the study period. Most assessments (17,816/26,646, 66.9%) were completed by female users. The mean user age was 34.3 years (SD 14.4 years), compared to a median age of 37.3 years of the general population. The most common initial symptom was abdominal pain (2060/26,646, 7.7%). A substantial number of assessments (12,357/26,646, 46.4%) were completed outside of typical physician office hours. Most users were advised to seek medical care on the same day (7299/26,646, 27.4%) or within 2-3 days (6301/26,646, 23.6%). Over a quarter of the assessments indicated a high degree of urgency (7723/26,646, 29.0%).

Conclusions: Users of the symptom checker chatbot were broadly representative of our patient population, although they skewed toward younger and female users. The triage recommendations were comparable to those of nurse-staffed telephone triage lines. Although the emergence of COVID-19 has increased the interest in remote medical assessment tools, it is important to take an evidence-based approach to their deployment.

(J Med Internet Res 2020;22(11):e20549) doi:10.2196/20549

KEYWORDS

symptom checker; chatbot; computer-assisted diagnosis; diagnostic self-evaluation; artificial intelligence; self-care; COVID-19

Introduction

Health care services in the United States are facing increasing levels of pressure, driven by a combination of aging populations, economic reform of health services, and more recently, the emergence of the COVID-19 pandemic [1].Training health care professionals is a slow process, and with widespread shortages of trained personnel and key vacancies throughout the system [2], scalable technological alternatives must be evaluated. One potential approach is a digital symptom checker, which is an artificial intelligence (AI)–supported software tool that uses a conversational "chatbot" format to ask questions about a patient's symptoms and returns a list of likely diagnoses to support self-diagnosis and appropriate triage [3].

The COVID-19 pandemic has brought new urgency to the consideration of chatbots due to the need to avoid face-to-face contact, preserve in-person care capacity, and triage patients at unprecedented volumes [4]. However, digital tools that impact care delivery should undergo rigorous evaluation that enables evidence-based determination of their efficacy. Symptom checker triage recommendations have been theorized to reduce unnecessary clinic and emergency room visits [5], and a recent study showed that completing a web-based symptom assessment reduced the urgency of the care that patients intended to seek [6]. However, little is known about aggregate triage recommendations generated by a symptom checker used in larger populations, and a number of recent reviews have called for more research to be shared [5,7,8]. Here, we describe the use characteristics and triage recommendations of one symptom checker chatbot deployed in partnership with a large, integrated health care system in Northern California.

Methods

Recruitment

The setting for this study is Sutter Health, a not-for-profit health care system in Northern California with 24 hospitals. In 2019, the symptom checker chatbot was introduced across the health system for broad use by any current and prospective patients over the age of 16 years. The chatbot was integrated into the main Sutter Health website (Figure 1) and web-based patient portal. Marketing was performed through several channels, including an email campaign to existing patients and social media advertisements.

Figure 1. Screenshot of the Sutter Health webpage during the symptom checker launch, May 2019.
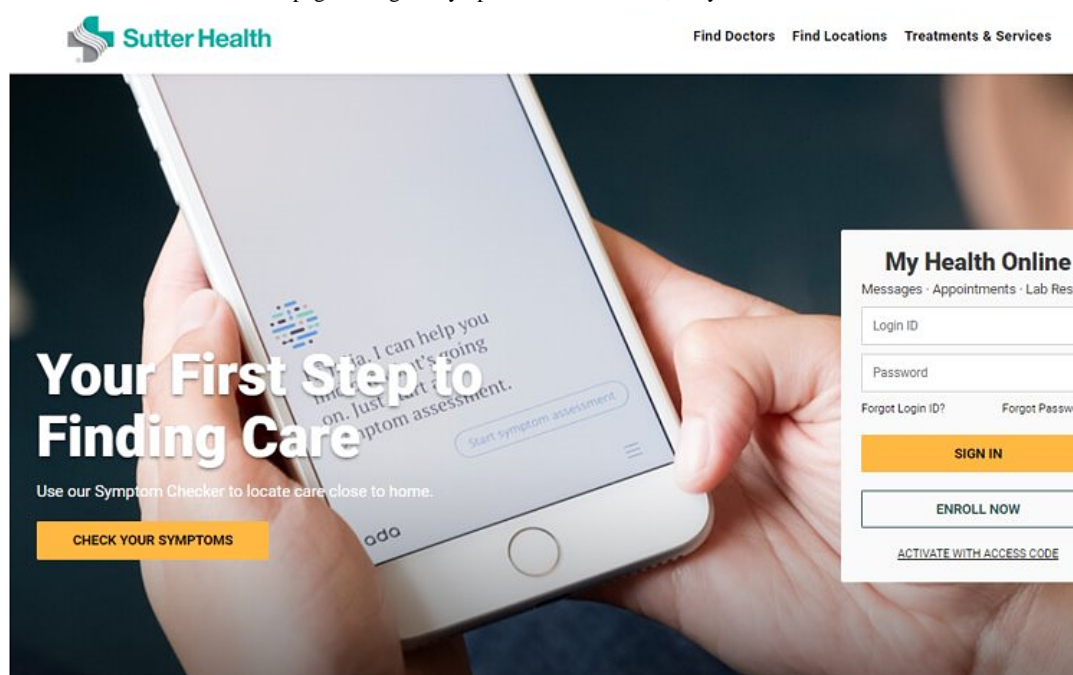
The data for this study encompass all symptom assessments completed from April 24, 2019, to February 1, 2020. This study was approved by the Sutter Health Institutional Review Board.

Symptom Checker

The symptom checker, developed by Ada Health (Ada Health, Berlin, Germany [9]), uses a conversational chatbot-style interface to elicit users' basic demographics and presenting symptoms as well as additional details such as symptom duration and severity. This information is analyzed by an AI algorithm to produce likely diagnoses and associated triage recommendations. The symptom checker assessments are anonymous; thus, recurrent users could not be identified, nor could use be linked to patient data within the electronic health record of the health system.

Statistical Analysis

Because this is a descriptive service improvement study, we had no falsifiable hypotheses; therefore, we did not undertake a formal power analysis. For comparison with our broader population, we extracted population-level demographics from the US Census Bureau data of Alameda County, one of the largest counties in Northern California served by Sutter Health [10]. Data analyzed included demographic information entered by the patient, initial symptoms reported, time of assessment, and triage advice generated by the symptom checker chatbot.

Triage advice took the form of one of eight possible suggestions, which were sorted into low, medium, or high acuity levels. Low acuity included suggestions to manage symptoms at home, seek medical advice in 2-3 weeks, or seek advice from a pharmacy. Medium acuity included suggestions to seek medical advice in 2-3 days or seek medical advice that same day. High acuity included suggestions to seek care within 4 hours, call an ambulance, or seek care in an emergency department.

Results

User Demographics and Time of Use

A total of 26,646 symptom assessments were completed during the study period, with no missing data. Most users (17,816/26,646, 66.9%) were female, and the remainder were male (8830/26,646, 33.1%). The comparator population of Alameda County is 50.9% female [10].

The mean age of the users was 34.3 years (SD 14.4 years); examination of subgroups (Table 1) revealed that the users were most commonly aged 30-39 years (7009/26,646, 26.3%). However, a sizable minority of users were in older age brackets; 3531/26,646 (13.3%) were aged 60 years or older. For comparison, the median age in Alameda County is 37.3 years, and 18.4% of the population is aged 60 or over [10].

Slightly less than half of the assessments (12,357/26,646, 46.4%) were completed outside of the typical physician office hours of 9 AM to 6 PM (Table 1). The most commonly reported initial symptom was abdominal pain (2060/26,646, 7.7%). The top 10 most commonly reported initial symptoms are shown in Table 2.

Table 1. Demographics and time of day of symptom checker use (N=26,646).

Characteristic	Count (%)
Gender	
Male	8830 (33.1)
Female	17,816 (66.9)
Age (years)	
<19	863 (3.2)
20-29	6441 (24.2)
30-39	7009 (26.3)
40-49	4663 (17.5)
50-59	4139 (15.5)
60-69	2209 (8.3)
70-79	951 (3.6)
80-89	247 (0.9)
90-99	44 (0.2)
>100	80 (0.3)
Time of assessment	
12 AM to 2:59 AM	1267 (4.8)
3 AM to 5:59 AM	1143 (4.3)
6 AM to 8:59 AM	3768 (14.1)
9 AM to 11:59 AM	5456 (20.5)
12 PM to 2:59 PM	4890 (18.4)
3 PM to 5:59 PM	3943 (14.8)
6 PM to 8:59 PM	3237 (12.2)
9 PM to 11:59 PM	2942 (11.0)

Table 2. The top 10 most common initial symptoms reported in the symptom checker (N=26,646).

Symptom	Count (%)
Abdominal pain	2060 (7.7)
Cough	1537 (5.8)
Headache	1085 (4.1)
Sore throat	897 (3.4)
Dizziness	621 (2.3)
Fatigue	559 (2.0)
Chest pain	534 (2.0)
Lower back pain	528 (2.0)
Diarrhea	466 (1.7)
Painful urination	460 (1.7)

Triage Urgency

Based on a user's symptom presentation, the symptom checker chatbot offered eight levels of triage advice, which were grouped into three levels of acuity (Table 3). Among the 26,646 assessments, 5323 (20.0%) directed the user to low acuity care, 13,600 (51.0%) directed the user to medium acuity care, and 7723 (29.0%) directed the user to high acuity care. The most common triage advice was to seek same-day medical care (Table 3).

Table 3. Advice and triage acuity levels of the assessments provided by the symptom checker (N=26,646).

Acuity level and advice	Count (%)
High	7723 (29.0)
Call an ambulance	1796 (6.7)
Seek emergency care	3703 (13.9)
Seek medical advice within 4 hours	2224 (8.3)
Medium	13,600 (51.0)
Seek medical advice within the same day	7299 (27.4)
Seek medical advice within 2-3 days	6301 (23.6)
Low	5323 (20.0)
Seek medical advice from a pharmacy	3433 (12.9)
Seek medical advice in 2-3 weeks	1617 (6.1)
Safely manage at home	273 (1.0)

Discussion

Principal Results

This study is one of the first published studies of the triage recommendations of an AI-driven symptom checker chatbot generated by a US-based patient population. Over a 9-month period, we saw robust use, particularly from younger and female users. Just under half of the assessments were completed outside of typical physician office hours, suggesting that there is a significant number of low-acuity concerns for which tailored guidance is not easily accessible during off-hours.

Understanding the user demographics of a symptom checker tool is an important milestone before subsequent, more nuanced questions can be answered. For example, there is a recognized need to study whether the use of symptom checkers augments patients' understanding and management of their illnesses, commonly described as "health literacy" [8]. Baseline health literacy, however, varies across patient demographics (including age) [11], and it must be taken into account when evaluating symptom checkers. Furthermore, health systems concerned about a widening "digital divide" driven by expanded virtual care options [12] rely on demographic information to identify and support patients who prefer to receive care through traditional channels.

This study is unique from previous work in that we assess the use of a symptom checker that has been deployed in partnership with a brick-and-mortar health system. Patient uncertainty about symptom checkers is recognized [13], and collaboration with a familiar health delivery mechanism could potentially improve patient engagement. To this end, our results show substantial use by older users (13.1% of users were aged 60 years and older), who are not typically considered to be heavy users of web-based tools. Furthermore, symptom checkers have been theorized to serve as surrogates for physician advice for patients who lack access to care [13]. Our population, however, is

predominately part of the Sutter Health care network; thus, they are using the symptom checker in conjunction with available in-person care options.

Comparison With Prior Work

A prior study of web-based symptom checkers found that users were predominantly female and had a mean age of 40 years [6]. This suggests that our symptom checker users are similar to users of other symptom checker tools, although skewing slightly younger. This younger age skew may be due to the promotion of the symptom checker chatbot by Sutter Health through predominately digital channels (eg, emails, website banners, digital newsletters), which may have created disproportionate awareness of the tool among patients who already use digital tools and are thus likely to be younger.

In previous studies, the triage recommendations of US-based nurse triage telephone lines reported high acuity recommendations in 19.7%-48.6% of calls versus 28.9% in the current study, medium acuity in 28%-48.2% of calls versus 50.9% in our study, and low acuity in 24%-36% of calls versus 20.1% in our study [14-16]. Accordingly, the distribution of the acuity of triage recommendations from the symptom checker chatbot appears to be generally comparable to that of US-based nurse triage telephone lines. These results are encouraging for the movement toward triage automation, which would enable the reallocation of clinicians to roles that better leverage their extensive training and would potentially improve health care staffing shortages.

Limitations

Limitations of this study include a potential lack of applicability to other symptom checkers, given that the results are wholly dependent on the configuration of a single symptom checker. The results were also influenced by the interest in digital health tools of a single geographic population; thus, they may not be nationally generalizable. True appropriateness of the triage recommendations cannot be assessed without patient-level comparisons against the existing gold standard of clinician-staffed triage telephone lines, and further research is needed to evaluate the diagnostic accuracy of the tool. Finally, although access to the symptom checker was provided through the Sutter Health webpage, use of the checker did not require any login or verification of affiliation with the Sutter Health care network; therefore, non-Sutter Health patients may have been included in our results. However, this use by non-Sutter Health patients is thought to be minimal because the Ada Health symptom checker chatbot is freely available elsewhere on the web, and navigating through the Sutter Health webpage requires additional steps that are not likely to be taken by the general population.

Conclusions

This study is one of the first published studies of the triage recommendations of an AI-driven symptom checker chatbot generated by a US-based patient population. Users of the chatbot were broadly representative of the general population of our region, although they skewed toward younger and female users. Our results suggest that the triage recommendations are acceptable; however, future research is needed to evaluate the medical accuracy of digital symptom assessment tools. While the recent emergence of COVID-19 and the need to take social distancing precautions may cause greater reliance on such tools, it is important to take an evidence-based approach to their deployment.

Conflicts of Interest

Authors VJ and AC are employees of Sutter Health.

References

1. Dieleman JL, Squires E, Bui AL, Campbell M, Chapin A, Hamavid H, et al. Factors Associated With Increases in US Health Care Spending, 1996-2013. JAMA 2017 Nov 07;318(17):1668-1678 [FREE Full text] [doi: 10.1001/jama.2017.15927] [Medline: 29114831]
2. Global strategy on human resources for health: Workforce 2030. World Health Organization. 2016. URL: https://www.who.int/hrh/resources/global_strategy_workforce2030_14_print.pdf?ua=1 [accessed 2020-05-04]
3. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
4. Espinoza J, Crown K, Kulkarni O. A Guide to Chatbots for COVID-19 Screening at Pediatric Health Care Facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
5. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
6. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of Use of Online Symptom Checkers With Patients' Plans for Seeking Care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]

XSL•FO

RenderX

7. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
8. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
9. Ada. URL: https://ada.com/ [accessed 2020-11-24]
10. Demographic Profile - California and Counties. State of California Employment Development Department. URL: https://www.labormarketinfo.edd.ca.gov/data/california-and-counties-demographic-profiles.html [accessed 2020-05-04]
11. Baker DW, Gazmararian JA, Sudano J, Patterson M. The association between age and health literacy among elderly persons. J Gerontol B Psychol Sci Soc Sci 2000 Nov 01;55(6):S368-S374. [doi: 10.1093/geronb/55.6.s368] [Medline: 11078114]
12. Mullangi S, Kaushal R, Ibrahim SA. Equity in the Age of Health Care Information Technology and Innovation. Med Care 2019;57:S106-S107. [doi: 10.1097/mlr.0000000000001033]
13. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient Perspectives on the Usefulness of an Artificial Intelligence-Assisted Symptom Checker: Cross-Sectional Survey Study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
14. Bogdan GM, Green JL, Swanson D, Gabow P, Dart RC. Evaluating patient compliance with nurse advice line recommendations and the impact on healthcare costs. Am J Manag Care 2004 Aug;10(8):534-542 [FREE Full text] [Medline: 15352529]
15. Moore JD, Saywell RM, Thakker N, Jones TA. An analysis of patient compliance with nurse recommendations from an after-hours call center. Am J Manag Care 2002 Apr;8(4):343-351 [FREE Full text] [Medline: 11950129]
16. O'Connell JM, Towles W, Yin M, Malakar CL. Patient Decision Making: Use of and Adherence to Telephone-Based Nurse Triage Recommendations. Med Decis Making 2016 Jun 25;22(4):309-317. [doi: 10.1177/0272989x0202200409]

Abbreviations

AI: artificial intelligence

Original Paper

Personalized Web-Based Weight Loss Behavior Change Program With and Without Dietitian Online Coaching for Adults With Overweight and Obesity: Randomized Controlled Trial

Alline Beleigoli[1,2,3*], MD, PhD; Andre Q Andrade[4*], MD, PhD; Maria De Fatima Diniz[3], MD, PhD; Antonio Luiz Ribeiro[5,6], MD, PhD

[1]Flinders Digital Health Research Centre, Flinders University, Adelaide, Australia

[2]Caring Futures Institute, Flinders University, Adelaide, Australia

[3]Post Graduation Course of Adult Health Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[4]Quality Use of Medicines and Pharmacy Research Centre, University of South Australia, Adelaide, Australia

[5]Department of Internal Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[6]Centre of Telehealth of the Hospital das Clinicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[*]these authors contributed equally

Corresponding Author:
Alline Beleigoli, MD, PhD
Flinders Digital Health Research Centre
Flinders University
Sturt North (N213)
GPO Box 2100
Adelaide, 5001
Australia
Phone: 61 8 82013303
Email: alline.beleigoli@flinders.edu.au

Abstract

Background: The effect of computer- or human-delivered personalized feedback on the effectivess of web-based behavior change platforms for weight loss is unclear.

Objective: We aimed to compare the effectiveness of a web-based behavior change intervention personalized through either computerized or human-delivered feedback with a nonpersonalized intervention in promoting weight loss in community-based adults with overweight or obesity.

Methods: This pragmatic, 3-group, parallel-arm, randomized trial recruited students and staff in a Brazilian public university who were aged 18 to 60 years, had a BMI of ≥25 kg/m², and were not pregnant. Participants were allocated to one of 3 groups: platform only (24-week behavior change program delivered using a web platform with personalized computer-delivered feedback), platform plus coaching (same 24-week web-based behavior change program plus 12 weeks of personalized feedback delivered online by a dietitian), or waiting list (nonpersonalized dietary and physical activity recommendations delivered through an e-booklet and videos). Self-reported weight at 24 weeks was the primary outcome. Changes in dietary and physical activity habits within 24 weeks were secondary outcomes.

Results: Among the 1298 participants, 375 (28.89%) were lost to follow-up. In the intention-to-treat analysis, the platform-only and platform plus coaching groups had greater mean weight loss than the waiting-list group at 24 weeks (–1.08 kg, 95% CI –1.41 to –0.75 vs –1.57 kg, 95% CI –1.92 to –1.22 vs –0.66 kg, 95% CI –0.98 to –0.34, respectively). The platform-only and platform plus coaching groups, compared with the waiting list group, had a greater increase in the consumption of vegetables (3%, 95% CI 1% to 6% vs 5%, 95% CI 2% to 8% vs –3%, 95% CI –5% to 0%) and fruits (9%, 95% CI 6% to 12% vs 6%, 95% CI 2% to 9% vs 2%, 95% CI 0% to 6%) and a larger reduction in ultraprocessed food intake (–18%, 95% CI –23% to –13% vs –25%, 95% CI –30% to –20% vs –12%, 95% CI –16% to –8%). Changes in physical activity did not differ across the groups. Engagement was higher in the platform plus coaching group than in the platform-only group (7.6 vs 5.2 completed sessions; *P*=.007). Longer usage of the platform was associated with clinically meaningful (≥5%) weight loss (odds ratio 1.02, 95% CI 1.01 to 1.04).

XSL•FO
RenderX

Conclusions: The web-based behavior change programs with computer- and human-delivered personalized feedback led to greater, albeit small-magnitude, weight loss within 24 weeks. Improvement in multiple dietary habits, but not physical activity, were also greater in the personalized programs compared with the nonpersonalized one. The human-delivered personalized feedback by the online dietitian coach increased user engagement with the program and was associated with a significantly higher chance of clinically meaningful weight loss.

Trial Registration: ClinicalTrials.gov NCT03435445; https://clinicaltrials.gov/ct2/show/NCT03435445

International Registered Report Identifier (IRRID): RR2-10.2196/10.1186/s12889-018-5882-y

KEYWORDS

obesity; overweight; healthy eating; diet; digital health; web platform; online coaching; personalized web interventions

Introduction

Obesity is associated with a range of health complications and might lead to increased mortality [1]. Interventions that target healthy diet and physical activity behaviors are the cornerstones of weight management. Despite their limited success, particularly for weight loss maintenance in the long term, these interventions remain pivotal due to their additional benefits, such as diabetes and premature mortality prevention [2].

The World Health Organization estimates that 39% of adults worldwide are overweight and 13% are obese [3]. This prevalence translates into 650 million adults with obesity worldwide, which means that reducing overweight and obesity are key public health challenges. Interventions across different levels—individual, interindividual (social support by family and close relationships), and environmental—are essential to tackle the obesity epidemic [4]. However, the delivery of individual interventions, such as weight counseling within primary care, faces several barriers. System capacity, lack of confidence and knowledge among health professionals, uncomfortable feelings among people living with excessive weight about discussing the issue, and the limited timely access to health professionals such as dietitians are some of the challenges the health system faces in tackling obesity at the individual level [5].

In this context, digital health, which is defined as the use of information and communication technologies for health improvement, and particularly web-based programs, have the potential to reach a large number of people and be widely accessible and cost-effective [6]. Affordability, anonymity, and opportunity are additional advantages of web-based weight loss programs in comparison with traditional face-to-face interventions [7]. Despite all these potential benefits, results have been heterogeneous in regard to weight loss results. In a recent systematic review, we found that behavior change interventions delivered exclusively through the web led to clinically small benefits in the short term and no significant long-term weight loss when compared with offline interventions in overweight and obese adults [8]. This seems to be related to low long-term adherence to web-delivered weight loss interventions, similar to the problems faced by face-to-face interventions.

Understanding the multiple dimensions of behavior may be the key to improving adherence to and impact of behavior change interventions. The Behavior Change Wheel model identifies individual capability, opportunity, and motivation as interconnected dimensions of behavior that should be addressed for change [9]. This model translates well into the concept of personalized applications, in which users' interactions with the application changes the experience and pathway to behavior change. In digital applications, personalization is usually enabled by algorithms but also by human-based guidance [10]. Health professional guidance, called "coaching" here, has been shown effective for behavior change in chronic obstructive pulmonary disease care [11] and heart failure improvement [12].

We aimed to investigate the impact of a personalized digital health behavior change intervention delivered exclusively via the web with and without online dietitian coaching on weight loss and on dietary and physical activity habits of people with overweight and obesity in the community compared with a minimal nonpersonalized intervention via the web. We also aimed to understand user engagement with the program.

Methods

Trial Design

The Online Platform for Healthy Weight Loss (POEmaS, from the abbreviation in Portuguese) study has been registered on ClinicalTrials.gov (NCT03435445), and the protocol with details of the intervention has been published elsewhere [13]. In brief, we conducted a 3-arm (1:1:1), parallel, randomized controlled trial, which recruited university students and staff in the Universidade Federal de Minas Gerais in Brazil. We used a pragmatic approach, with enrollment and outcomes assessment being exclusively online.

Participants and Recruitment

University students and staff were invited through banners, posters, and mass media emails from September 25, 2017, to October 24, 2017. Participants were instructed to access a website, where they were informed about the aims of the study, inclusion criteria (aged 18 to 60 years, BMI ≥ 25 kg/m^2, intention to lose weight through a behavior change program, and web access), and exclusion criteria (pregnancy, participation in any other weight loss program, or presence of conditions that demand specific dietary or physical activity recommendations, such as diabetes, heart failure, coronary artery disease, kidney disease, hepatic disease, cancer, phenylketonuria, celiac disease, food allergies, or bariatric surgery history).

Randomization and Allocation

Those who were eligible were allocated to one of 3 study groups using a stratified randomized block design by sex and category of body mass index (25 to <30 or ≥30 kg/m^2) using blocks of variable length (either 3 or 6). Then, participants received an email with information about the activities available to the group they were allocated to. The random allocation sequence and the algorithm for randomization were developed by a team of information technology specialists that did not participate in the recruitment or assessment processes. Those who did not complete the questionnaires about dietary and physical activity habits during the onboarding process could not proceed to the use of the platform.

Study Groups

A detailed description of the rationale for the development of the intervention can be found elsewhere [14]. The waiting-list (control) group received a nonpersonalized minimal intervention based on dietary and physical activity recommendations delivered through a downloadable e-booklet and four 5-minute videos with information about health consequences of obesity, healthy dietary recommendations, healthy physical activity recommendations, and daily life strategies for the adoption of healthy behaviors. These resources were available to participants of this group through the platform from the beginning of the trial and could be accessed at any time. Moreover, like the intervention groups, this group received emails reminding them to report their weight and habits through the platform at 12 and 24 weeks after the trial baseline. By the end of the trial, these participants gained access to the weight loss program delivered through the web platform. This platform was adapted from a commercial software that has been used for multiple workforce behavior change and wellness interventions in Brazil.

The platform-only group was given access to a weight loss program delivered through the web-based platform. The program was based on diet [15] and physical activity [16] guidelines and on the Behavior Change Wheel model [9]. It comprised a total of 24 weekly sessions (12 weeks of an intensive program and 12 weeks of a maintenance program). The behavior change techniques (BCTs) [17] that were applied to address the capability, opportunity, and motivation of the participants in this group compared with the other groups can be seen in Multimedia Appendix 1. These BCTs were delivered using a range of software functionalities, such as short educational readings and videos, graphical and interactional tools, qualitative and quantitative (food diary) dietary monitoring, physical activity self-monitoring tasks, interactive games that created opportunities to invite friends and adopt healthy habits in daily life, and an online social network embedded in the platform and moderated by physicians and dietitians. Personalized feedback generated by a computational algorithm that took into account the goals set by each participant and the data on habits reported by the participant in initial questionnaires and through the self-monitoring tools was provided to participants from the fourth week of the intervention. This personalized feedback comprised feedback on behaviors and suggestions of strategies to improve their success in accordance with their individual goals. Furthermore, the platform suggested different modes of interaction (texts, social interaction, challenges) according to patterns of use during the first 4 weeks.

The platform and coaching group followed the same 24-week weight loss program delivered by the platform plus a 12-week initial course of online personalized education and feedback by a dietitian. The interactions between the participant and the dietitian could be initiated by either side through a private forum embedded in the platform. There was no limit to the amount of contact between them.

Although all groups received similar information about the target behaviors, there were substantive differences between the interventions received by the waiting-list group and the platform groups regarding the mode of delivery of the information and promotion of behavior change (Multimedia Appendix 1). An example of how the platform delivered the behavior change techniques can be demonstrated by the target behavior of increasing the intake of vegetables. The control group received information on the health consequences of this habit and instructions on the how to adopt this behavior through videos and recipes in the e-booklet. The capability of the groups using the platform was addressed through similar recommendations with short texts and videos. In addition, to address opportunity, these groups also received vegetable-rich recipes through the platform around the time of their main meals and were given challenges to post photos of vegetable-rich meals on the social media network. To enhance their motivation for this specific behavior, they scored points on their health score each time they reported vegetable intake on the data input tools on the platform or when they posted a photo in reference to a related challenge. An algorithm enabled them to receive tailored messages of feedback on that specific behavior based on the data they had input on the platform over the previous 4 weeks. Suggestions of resources available on the platform that could help them achieve their goals were also part of the feedback. For the platform plus coaching group, this process was enhanced by personalized feedback from the dietitian through a private chat forum.

The feedback provided by the coach included a review about the participant's goal setting for behaviors and outcomes. Moreover, the dietitian specifically promoted self-monitoring of behaviors and outcomes, as well as emotional social support. When the participant had completed a food diary, the dietitian provided individualized feedback and informative social support regarding dietary quality and quantity. Reflective motivation was addressed by the development of action plans and problem-solving strategies pertinent to the individual's circumstances. Prescripted responses to common topics were also used by the dietitian for feedback.

Outcome Measures

We adopted a pragmatic approach by considering a real-world telehealth context and used self-reported (rather than measured) weight and BMI changes as the primary outcomes of the study. Weight reporting was required during baseline, and participants were encouraged through messages on the platform to continuously report it for the duration of the study. A validation study was conducted with a random sample of 12.5% of the study population to investigate the agreement between

anthropometric measures that were self-reported by the participants and those that were measured by a trained research team using standardized and validated methods. Differences between self-reported and measured weight and BMI were clinically small and statistically nonsignificant, which led to high agreement between the self-reported weight and BMI and the measured anthropometry [18].

As secondary outcomes, the number of daily vegetable and fruit portions and the weekly consumption of sweetened beverages and ultraprocessed foods that were reported over the platform after 24 weeks were assessed through the Brazilian food frequency questionnaire (*questionario de frequência alimentar*) [19]. Moderate and vigorous physical activity was assessed by the Brief Physical Activity Assessment Questionnaire [20].

Outcomes assessors were blinded to group allocation.

Sample Size

Based on 90% power to detect a significant difference of a 4 kg weight loss between groups, assuming the SD of weight would be 6.0 and using a 2-sided significance level of .05 and a 40% attrition rate, a sample size of 90 participants was calculated for each group [21].

Data Analysis

Intention-to-treat analysis was performed for the primary outcomes at 12 and 24 weeks and for the secondary outcomes at 24 weeks. Analysis of covariance was used to test for differences in weight and BMI loss between groups at each time point, adjusted for treatment group as the predictor variable of interest and weight at baseline as a covariate. Statistical significance of the primary efficacy analysis (at 12 and 24 weeks) was adjusted for multiple testing procedures (Bonferroni). Analyses of the secondary outcomes were set at 24 weeks and used a 2-sided .05 significance level. Sensitivity analysis according to BMI (25-29.9 kg/m^2 and ≥30 kg/m^2) was performed.

Although not planned in the research protocol, due to the high number of missing values for the primary outcome, we performed multiple imputation by fitting logistic and linear regression models with both the predictors and the outcome as well as with other variables regarded as important to explain the missing values [19]. This procedure generated 5 complete data sets, which were used to estimate the association between group allocation and primary and secondary outcomes.

We compared clinically meaningful weight loss, defined as a ≥5% loss, weight stability (−5% to 5% difference), and weight gain (≥5% gain) across groups using chi-square tests. To investigate the association between adherence and clinically meaningful weight loss, we performed binary logistic regression with weight loss ≥5% (no or yes) at 24 weeks as the response variable and number of accesses to the platform, group (platform or platform plus coach), initial weight, and gender as covariates.

Data preprocessing and statistical analysis were done using the Python packages Pandas [22] and SciPy [23]. Multiple imputation was performed using IBM SPSS (version 18; IBM Corp).

Ethics

The study was approved by the Ethics Committee of the Universidade Federal de Minas Gerais (CAAE: 73545717.5.0000.5149). All participants signed an online informed consent form prior to recruitment.

Results

Participant Characteristics

A total of 3745 participants were assessed for eligibility, and 1298 were allocated to one of the 3 arms. This number is considerably higher than the sample size calculated for the study (n=270). We attribute this to a very successful recruitment process, which involved a mass communication strategy. Knowing from previous studies that interventions for weight loss, including web-based interventions, are usually associated with high dropout rates [8], we decided to increase our team capacity and resources to follow up with this higher number of participants. Across all groups, 375 of the 1298 participants (28.89%) were lost to follow-up. Participant workflow can be seen in Figure 1 and participants' characteristics at baseline (N=1298) can be seen in Table 1.

Figure 1. CONSORT flow diagram for the POEmaS randomized controlled trial. CONSORT: Consolidated Standards of Reporting Trials; POEmaS: Online Platform for Healthy Weight Loss.
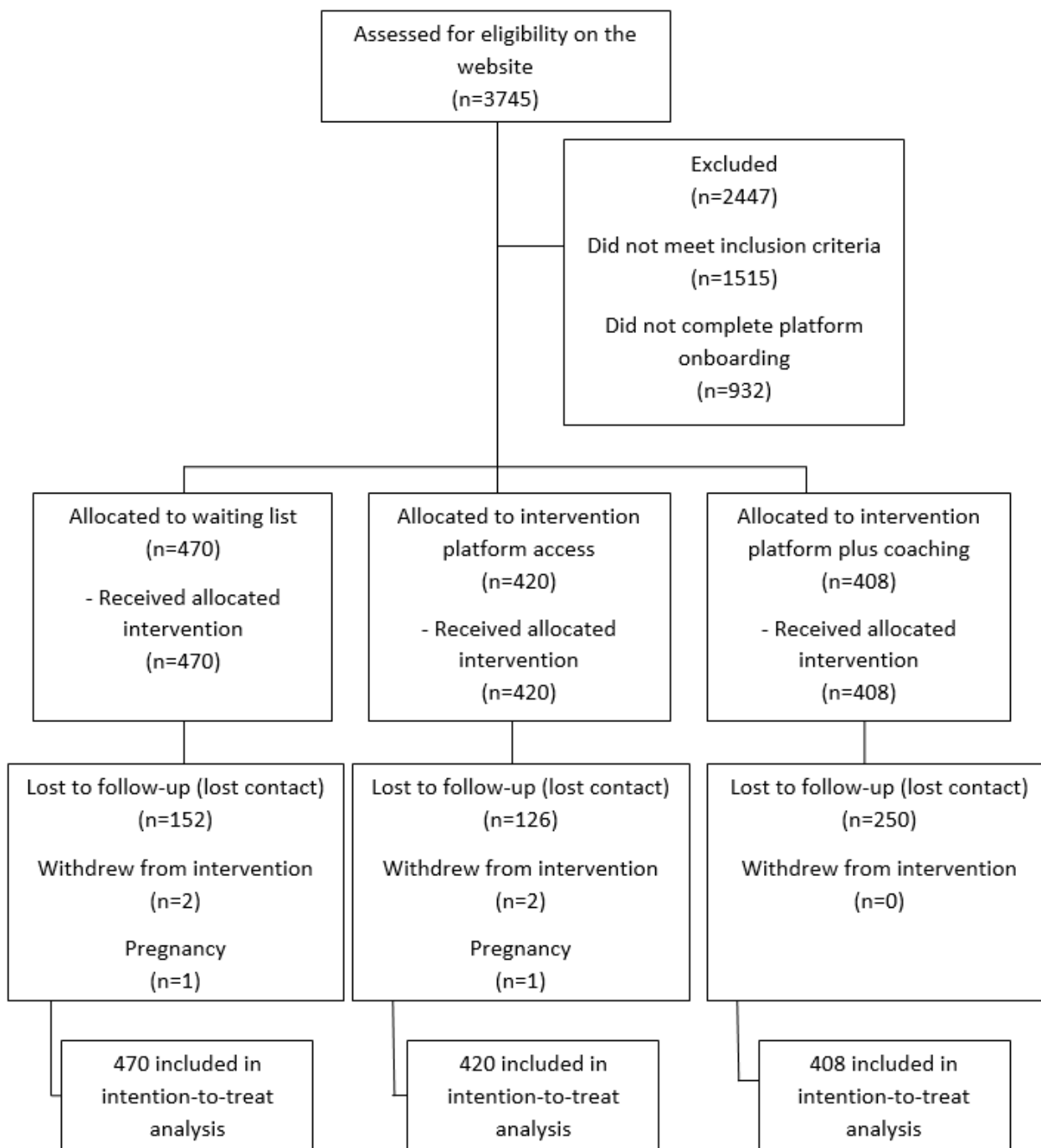
Table 1. Participants' characteristics at baseline.

Characteristic	Total (n=1298)	Waiting list (control) (n=470)	Platform only (n=420)	Platform and coaching (n=408)
Weight (kg), mean (95% CI)	82.8 (81.9-83.6)	82.6 (81.3-84.0)	83.4 (81.7-85.0)	82.3 (80.8-83.7)
Age (years), mean (95% CI)	33.6 (33.0-34.2)	33.4 (32.4-34.4)	34.4 (33.4-35.6)	33.0 (31.9-34.0)
BMI (kg/m^2), mean (95% CI)	29.89 (29.66-30.13)	29.73 (29.37-30.08)	30.12 (29.67-30.58)	29.85 (29.44-30.26)
Female, n (%)	996 (76.7)	362 (77.0)	315 (75.0)	319 (78.2)
Vegetable intake[a], mean (95% CI)	3.1 (3.1-3.2)	3.1 (3.0-3.2)	3.1 (3.0-3.2)	3.1 (3.0-3.2)
Fruit intake[a], mean (95% CI)	2.8 (2.7-2.8)	2.8 (2.7-2.9)	2.7 (2.6-2.8)	2.8 (2.7-2.9)
Whole grains intake[a], mean (95% CI)	1.8 (1.7-1.8)	1.7 (1.6-1.9)	1.7 (1.6-1.9)	1.8 (1.7-1.8)
Ultraprocessed foods[a], mean (95% CI)	2.7 (2.7-2.8)	2.8 (2.7-2.9)	2.7 (2.6-2.8)	2.8 (2.6-2.8)
Sweetened beverages[a], mean (95% CI)	1.8 (1.7-1.9)	1.8 (1.7-1.9)	1.8 (1.7-1.9)	1.8 (1.7-1.9)
Moderate physical activity[b], mean (95% CI)	2.4 (2.2-2.6)	2.6 (2.4-2.8)	2.9 (2.5-3.0)	1.8 (1.6-2.0)
Vigorous physical activity[b], mean (95% CI)	1.3 (1.1-1.8)	1.5 (1.1-1.8)	1.6 (1.3-2.0)	0.6 (0.4-1.3)
Stages of change for physical activity[c], n (%)				
Precontemplation	46 (3.9)	19 (4.0)	12 (2.9)	15 (3.7)
Contemplation	435 (36.7)	146 (31.1)	163 (38.8)	126 (30.9)
Preparation	276 (23.4)	86 (18.3)	96 (22.9)	94 (23.0)
Action	256 (21.7)	86 (18.3)	79 (18.8)	91 (22.3)
Maintenance	169 (14.3)	81 (17.2)	45 (10.7)	43 (10.5)

[a]Measured in servings per day.

[b]Measured as days per week exercising for more than 10 minutes.

[c]Precontemplation=not intending to engage in physical activity within 6 months; contemplation=intending to engage in physical activity within 6 months; preparation=intending to engage in physical activity within 30 days; action=physically active for less than 6 months; maintenance=physically active for more than 6 months.

Primary Outcomes

The absolute weight loss and BMI loss at 12 weeks were higher in the platform groups than in the waiting-list group, and there was no difference between the intervention groups (Table 2). At 24 weeks, weight loss and BMI loss were superior in the platform plus coaching group in comparison with the waiting-list group. A minimum 5% weight loss occurred more frequently in the platform-only (83/420, 19.8%) and platform plus coaching (64/408, 15.7%) groups than in the waiting-list group (61/270, 13.0%; *P*=.001), as seen in Figure 2. These results did not change when participants with overweight and obesity were analyzed separately (Multimedia Appendices 2 and 3) or when analysis included only participants with two or more weight reports (ie, no multiple imputation performed) (Multimedia Appendix 4).

Table 2. Primary outcomes after 12 and 24 weeks of follow-up according to intention-to-treat analysis.

Outcomes	Waiting list, mean (95% CI) (n=470)	Platform only, mean (95% CI) (n=420)	Platform plus coaching, mean (95% CI) (n=408)	P value[a]
12 weeks				
Weight (kg)	82.06 (80.71 to 83.42)	82.23 (80.58 to 83.89)	80.88 (79.42 to 83.35)	N/A[b]
Weight change (kg)	–0.56 (–0.83 to –0.30)	–1.14 (–1.42 to –0.85)	–1.36 (–1.65 to –0.80)	<.001
BMI (kg/m^2)	29.52 (29.15 to 29.89)	29.71 (29.25 to 30.18)	29.36 (28.93 to 29.78)	N/A
BMI change (kg/m^2)	–0.20 (–0.30 to –0.11)	–0.41 (–0.51 to –0.31)	–0.50 (–0.60 to –0.39)	<.001
24 weeks				
Weight (kg)	81.97 (80.62 to 83.32)	82.29 (80.66 to 83.92)	80.68 (79.22 to 82.15)	N/A
Weight change (kg)	–0.66 (–0.98 to –0.34)	–1.08 (–1.41 to –0.75)	–1.57 (–1.92 to –1.22)	.001
BMI (kg/m^2)	29.49 (29.12 to 29.86)	29.74 (29.29 to 30.20)	29.29 (28.86 to 29.72)	N/A
BMI change (kg/m^2)	–0.24 (–0.35 to –0.12)	–0.38 (–0.50 to –0.26)	–0.56 (–0.69 to –0.43)	.001

[a]P values based on comparisons across the 3 groups by analysis of covariance. For weight change at 12 weeks, P value for comparison between groups A (waiting list) and B (platform only) was .01, between groups A and C (platform plus coaching) was <.001, and between groups B and C was .80. For weight change at 24 weeks, P value for comparison between groups A and B was .23, between groups A and C was <.001, and between groups B and C was .14. For BMI change at 12 weeks, P value for comparison between groups A and B was .01, between groups A and C was <.001, and between groups B and C was .75. For BMI change at 24 weeks, P value for comparison between groups A and B was .28, between groups A and C was .001, and between groups B and C was .80.

[b]N/A: not applicable

Figure 2. Clinically meaningful weight loss, stability, and gain per group at 24 weeks.
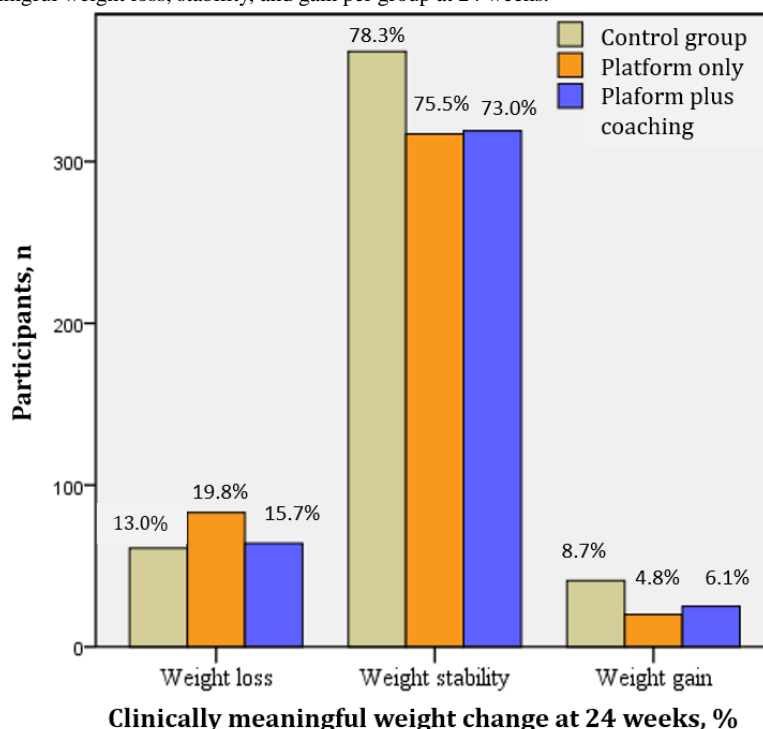
Secondary Outcomes

Both the platform-only and the platform plus coaching group had a greater increase in vegetable and fruit intake and a greater reduction in ultraprocessed food intake at 24 weeks in comparison with the control group. The reduction in sweetened beverage consumption was higher in the platform plus coaching group than in the platform-only group. Changes in other dietary habits and in moderate and physical activity duration were not different across groups (Table 3).

Table 3. Dietary and physical activity habits at 24 weeks and percent change from baseline (95% CI) across study groups.

Outcomes	Waiting list (control), mean (95% CI)	Platform only, mean (95% CI)	Platform plus coaching, mean (95% CI)	P value[a]
Vegetable intake (servings/day)	3.1 (3.0 to 3.2)	3.2 (3.1 to 3.3)	3.3 (3.2 to 3.4)	N/A[b]
Vegetable intake change[c]	–3 (–5 to 0)	3.0 (1 to 6)	5 (2 to 8)	.001
Fruit intake (servings/day)	2.9 (2.8 to 2.9)	3.0 (2.9 to 3.0)	2.9 (2.9 to 3.0)	N/A
Fruit intake change[c]	2 (–0 to 6)	9 (6 to 12)	6 (2 to 9)	.02
Whole grains (servings/day)	1.8 (1.7 to 1.9)	1.8 (1.7 to 1.9)	1.7 (1.6 to 1.8)	N/A
Whole grains change[c]	3 (–2 to 9)	2 (–1 to 10)	–5 (–12 to 2)	.11
Ultraprocessed foods (servings/day)	2.5 (2.4 to 2.6)	2.3 (2.2 to 2.4)	2.2 (2.1 to 2.3)	N/A
Ultraprocessed foods change[c]	–12 (–16 to –8)	–18 (–23 to –13)	–25 (–30 to –20)	.005
Sweetened beverages (servings/day)	1.7 (1.6 to 1.8)	1.8 (1.7 to 1.9)	1.6 (1.5 to 1.7)	N/A
Sweetened beverages change[c]	–6 (–12 to 0)	0 (–5 to 7)	–14 (–21 to –8)	.008
Moderate activity duration[d]	2.4 (2.2 to 2.6)	2.7 (2.4 to 2.9)	2.3 (2.1 to 2.5)	N/A
Moderate activity duration change[e]	–15 (–25 to –3)	–4 (–13 to –5)	23 (9 to 37)	.21
Vigorous activity duration[d]	1.3 (0.9 to 1.8)	1.6 (1.1 to 2.2)	1.1 (0.6 to 1.5)	N/A
Vigorous activity duration change[e]	–14 (–28 to 0)	2 (–9 to 13)	4 (–18 to 11)	.19

[a]P value based on analysis of covariance. For vegetable intake change, difference between groups A (waiting list) and B (platform only) was .03, between groups A and C (platform plus coaching) was .001, and between groups B and C was .71. For fruit intake change, difference between groups A and B was .01, between groups A and C was .49, and between groups B and C was .47. For ultraprocessed food intake change, difference between groups A and B was .35, between groups A and C was .003, and between groups B and C was .28. For sweetened beverage intake change, difference between groups A and B was .35, between groups A and C was .26, and between groups B and C was .01.

[b]N/A: not applicable.

[c]Percent change in servings per day from baseline.

[d]Measured as days per week exercising for more than 10 minutes.

[e]Percent change in number of days per week from baseline.

Engagement

The mean number of sessions completed was 5.2 (95% CI 4.1-6.3) and 7.6 (95% CI 6.0-9.1) for the platform-only group and the platform plus coaching group, respectively (P=.007). Participants' interactions with the platform showed considerable initial attrition, which can be verified by the large number of participants who completed only 1 session: 126 of 420 (30.0%) in the platform-only and 97 of 408 (23.7%) in the platform plus coaching group. The number of participants who completed sessions at 12 and 24 weeks across all groups is shown in Figure 3. Except for whole grain intake, which was higher among completers, baseline characteristics of completers and noncompleters did not differ (Multimedia Appendix 5).

Functionalities that delivered self-monitoring of behavior and social support were the most accessed ones in both groups, with a higher mean access rate of the latter by the platform plus coaching group (48.7 times, 95% CI 37.8-59.6 vs 32.5 times, 95% CI 25.0-39.9; P=.02). The use of each platform functionality and the corresponding behavior change techniques delivered by the functionality are displayed in Table 4.

The total number of sessions completed by the participants was independently associated with clinically significant weight loss (≥5%) at 24 weeks (odds ratio 1.02, 95% CI 1.01-1.04) when adjusted for initial weight, study group (platform or platform plus coaching), and gender.

Figure 3. Number of participants completing sessions per group (group B is the platform-only group and group C is the platform plus coaching group) over 24 weeks.
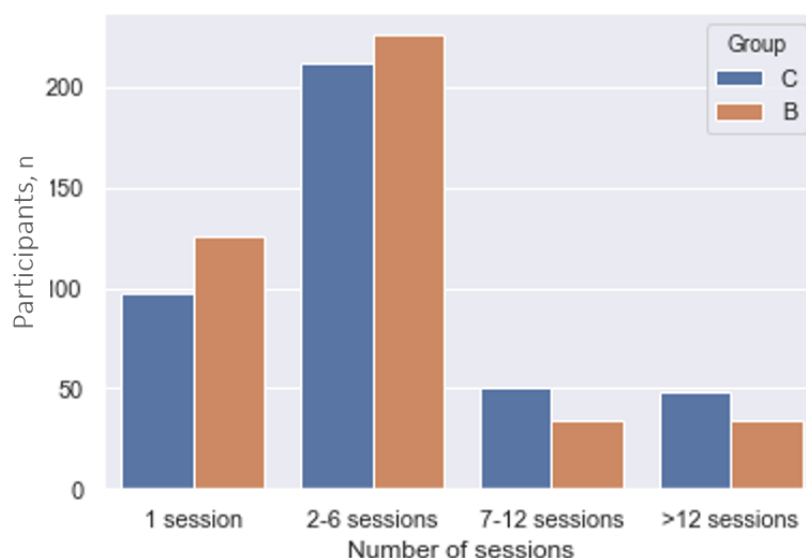
Table 4. Number of accesses to platform functionalities and the corresponding behavior change techniques between the platform-only and platform plus coaching groups.

Platform functionality (BCTs[a])	Platform only, mean (95% CI)	Platform plus coaching, mean (95% CI)	*P* value
Weight report (self-monitoring of outcome of behavior)	2.7 (2.3-3.1)	3.3 (2.5-4.1)	.21
Behavior report (self-monitoring of behavior)	49.7 (27.1-72.3)	72.1 (43.5-100.6)	.23
Profile (goal setting of outcome; review outcome goals; feedback on behavior)	7.1 (5.6-8.5)	7.8 (6.2-9.5)	.49
Small texts (action planning; instruction on how to perform the behavior; information about antecedents, health consequences, and emotional consequences; reduce negative emotions; verbal persuasion about capability; restructuring the social environment)	5.8 (4.9-6.7)	5.9 (5.0-6.8)	.85
Challenges and gamification resources (problem solving; restructuring the social environment; avoidance and reducing exposure to cues for the behavior; imaginary reward)	0.2 (0.1-0.2)	0.2 (0.1-0.2)	.91
Online social network (unspecified, practical, and emotional social support; avoidance and reducing exposure to cues for the behavior)	32.5 (25.0-39.9)	48.7 (37.8-59.6)	.02

[a]BCT: behavior change technique.

Discussion

Primary and Secondary Outcomes

Weight and BMI loss were greater after 12 weeks and 24 weeks in the groups using the platform (with or without coaching) than in the group receiving a minimal intervention. The magnitude of weight loss (<2 kg) across the groups was small, which is similar to mean differences found in recent meta-analyses, in which weight loss programs delivered through the web were compared with face-to-face or to no interventions for people with overweight or obesity [8,24]. However, clinically meaningful weight loss (≥5%) was significantly more common in the platform groups with personalized feedback (83/420, 19.8% of the platform-only group and 64/408, 15.7% of the platform plus coaching group) than in the control group that received a nonpersonalized intervention (61/470, 13.0%). Moreover, compared with the control group, ultraprocessed food and sweetened beverage consumption in the platform groups decreased, while vegetable and fruit intake increased. This is a very positive result because only about 34.7% of Brazilian adults consume 5 or more servings per day of vegetables and fruits [25]. The importance of these results is

also related to the fact that health benefits associated with vegetable and fruit consumption are independent of weight loss [26]. The superiority of the platform groups compared with the control group for both primary and secondary outcomes suggests that different behavior change techniques that address capability, opportunity, and motivation have a greater effect on promoting dietary behavior change and weight loss than a nonpersonalized focus on capability only [27].

Although our study design cannot point to which BCTs explain the difference in effectiveness across the groups, the BCTs of social support, personalized feedback, and self-monitoring of behaviors, which only the platform groups received, were important intervention differences. The efficacy of these BCTs in promoting weight loss has been reported by other studies [28].

Similar to other studies, there was no difference in short-term weight loss [29] and in changes in diet and physical activity between the group that received computer-based personalized feedback (platform only) and the group that also received the human-delivered personalized feedback (platform plus coaching). Despite this lack of differences in the outcomes, the addition of a health professional coaching service increased platform usage. This suggests that the feeling of having a human factor [10,30] or of being supervised [31] increases engagement, which might be particularly important for long-term weight maintenance [29].

Engagement

User engagement results were similar to other large-scale weight loss interventions via the web [8]. Most of the losses to follow-up occurred at the beginning of the intervention across all groups, particularly in the waiting-list group (152/470, 32.3%). The study design mimicked real-world recruitment and usage conditions. In this sense, participants received email reminders to report their weight but not to use the platform, and there were no financial incentives or contact between the research team and the participants. The broad recruitment strategy led to the enrollment of a large proportion of individuals who were probably not predisposed to engage, as suggested by the high proportion of individuals (481/1298, 37.06% across all groups) who reported being on the precontemplation or contemplation stage of change for physical activity, according to the transtheoretical model–based questionnaire given to all participants at baseline (Table 1). A recruitment based on the stages of readiness to change for weight loss might have yielded different results [32]. Moreover, the mandatory completion of questionnaires in the dashboard before being able to use the platform and the technical issues in the beginning of the intervention (despite being promptly corrected) might have contributed to the loss of participants who were not highly motivated.

Our results also suggest that adherence to the behavior change intervention is key to the weight loss outcome. Our analysis showed that each additional session completed by participants was associated with a 2% increase (95% CI 1%-4%) in the chance of achieving clinically significant weight loss at 24 weeks. This demonstrates that engagement is a key factor for the success of online interventions, as reported by other studies, including those for long-term weight maintenance [10,33,34].

Implications to Practice

The intervention delivery package, which involved broad recruitment and few follow-up visit requirements, mimics the conditions of a low-demand telehealth intervention in communities (rather than for patients in specific health care settings). In this context, a behavior change program delivered through a web platform might be an effective solution for public health interventions that promote weight loss and increased fruit and vegetable intake in the short term.

The inclusion of a human-based personalization strategy using an online dietitian coach did not appear to be more effective than the computer-based personalization strategy after 6 months of follow-up. This suggests that web platforms enhanced by algorithm-generated personalized feedback might be a good strategy for tackling overweight and obesity through lifestyle habit changes in the short term. This can be particularly useful in contexts in which the demand for professionals to support people with excessive weight cannot be met. However, since the human-delivered feedback led to higher rates of engagement and longer use of the platform, the human-delivered feedback strategy might be useful for participants with a high risk of abandoning the program, such as those reporting low levels of preintervention motivation or with multiple weight loss attempts [31].

Strengths and Limitations

The large scale of this clinical trial (1298 participants) and its pragmatic nature are main strengths of this study. To our best knowledge, this is the largest trial evaluating health professional coaching for behavior change in a large and diverse group of participants recruited in the community. Such scale was only possible with an open recruitment strategy, which reduces the barrier for enrollment but also brings some limitations. The first limitation is related to not being able to take standardized and repeated measurements of weight and BMI, since we had no direct contact with most participants. Additionally, the large number of noncompleters, which is common to web-based interventions, increased the risk of bias in our results.

Conclusion

A behavior change program for weight loss delivered through a web-based platform led to greater weight loss, increased fruit and vegetable intake, and reduced ultraprocessed food consumption compared with a minimal intervention, as measured up to 6 months into the intervention. The platform enhanced by human-delivered personalized feedback was not superior to the platform with a computer-based personalized approach for weight loss. However, it led to higher levels of engagement, which were associated, albeit weakly, with higher odds of achieving clinically significant weight loss.

Acknowledgments

Authors' Contributions

ALR and AB obtained the funding and coordinated the study. AB, AQA, MDFD, and ALR designed the study. AB and MDFD advised on the development of the behavior change intervention. AB and MDFD validated the study. AQA and AB analyzed the data. AB and AQA wrote the first substantial draft of the paper, and AB is the guarantor. All authors read, revised, and approved the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

AQA was the owner of the software company that provided the web platform by the time the study was conducted. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Behavior change techniques (BCTs) across the study groups.
[DOCX File , 9 KB - jmir_v22i11e17494_app1.docx]

Multimedia Appendix 2
Primary outcomes analysis for participants with overweight.
[DOCX File , 10 KB - jmir_v22i11e17494_app2.docx]

Multimedia Appendix 3
Primary outcomes analysis for participants with obesity.
[DOCX File , 10 KB - jmir_v22i11e17494_app3.docx]

Multimedia Appendix 4
Primary outcomes at 12 and 24 weeks for participants with more than one report of weight.
[DOCX File , 13 KB - jmir_v22i11e17494_app4.docx]

Multimedia Appendix 5
Comparison of baseline characteristics of participants who completed and those who did not complete the intervention.
[DOCX File , 13 KB - jmir_v22i11e17494_app5.docx]

References

1.   GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N Engl J Med 2017 Jul 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]

2.   Ma C, Avenell A, Bolland M, Hudson J, Stewart F, Robertson C, et al. Effects of weight loss interventions for adults who are obese on mortality, cardiovascular disease, and cancer: systematic review and meta-analysis. BMJ 2017 Nov 14;359:j4849. [doi: 10.1136/bmj.j4849] [Medline: 29138133]

3.   Obesity and overweight. World Health Organization. URL: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-10-01]

4.   4 P. A. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/obesity/health_equity/addressingtheissue.html [accessed 2020-10-08]

5.   Blackburn M, Stathi A, Keogh E, Eccleston C. Raising the topic of weight in general practice: perspectives of GPs and primary care nurses. BMJ Open 2015 Aug 07;5(8):e008546. [doi: 10.1136/bmjopen-2015-008546] [Medline: 26254471]

6.   Classification of digital health issues v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. URL: http://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2020-07-01]

XSL•FO

RenderX

7. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto SAP, Meira W, et al. Factors Associated With Weight Change in Online Weight Management Communities: A Case Study in the LoseIt Reddit Community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]

8. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MDFH, Ribeiro AL. Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]

9. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

10. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

11. Jolly K, Sidhu MS, Hewitt CA, Coventry PA, Daley A, Jordan R, et al. Self management of patients with mild COPD in primary care: randomised controlled trial. BMJ 2018 Jun 13;361:k2241. [doi: 10.1136/bmj.k2241] [Medline: 29899047]

12. Tiede M, Dwinger S, Herbarth L, Härter M, Dirmaier J. Long-term effectiveness of telephone-based health coaching for heart failure patients: A post-only randomised controlled trial. J Telemed Telecare 2017 Sep;23(8):716-724. [doi: 10.1177/1357633X16668436] [Medline: 27605214]

13. Beleigoli AM, Queiroz de Andrade A, Haueisen Diniz MDF, Alvares RS, Ribeiro AL. Online platform for healthy weight loss in adults with overweight and obesity - the "POEmaS" project: a randomized controlled trial. BMC Public Health 2018 Aug 01;18(1):945 [FREE Full text] [doi: 10.1186/s12889-018-5882-y] [Medline: 30068393]

14. Beleigoli AM, Andrade AQ, Diniz MFH, Alvares RS, Ferreira MH, Silva LA, et al. Using the Behaviour Change Wheel for Designing an Online Platform for Healthy Weight Loss - "POEmaS". Stud Health Technol Inform 2018;254:1-10. [Medline: 30306951]

15. Dietary Guidelines for the Brazilian population. Ministry of Health of Brazil. URL: http://bvsms.saude.gov.br/bvs/publicacoes/dietary_guidelines_brazilian_population.pdf [accessed 2020-10-10]

16. US Preventive Services Task Force, Grossman DC, Bibbins-Domingo K, Curry SJ, Barry MJ, Davidson KW, et al. Behavioral Counseling to Promote a Healthful Diet and Physical Activity for Cardiovascular Disease Prevention in Adults Without Cardiovascular Risk Factors: US Preventive Services Task Force Recommendation Statement. JAMA 2017 Jul 11;318(2):167-174. [doi: 10.1001/jama.2017.7171] [Medline: 28697260]

17. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

18. Beleigoli A, Andrade A, Diniz M, Alvares R, Ferreira M, Silva L, et al. Validation of Anthropometric Measures Self-Reported in a Randomized Controlled Trial of a Web-Based Platform for Weight Loss. Stud Health Technol Inform 2019 Aug 08;266:30-36. [doi: 10.3233/SHTI190769] [Medline: 31397298]

19. Molina MDCB, Benseñor IM, Cardoso LDO, Velasquez-Melendez G, Drehmer M, Pereira TSS, et al. Reprodutibilidade e validade relativa do Questionário de Frequência Alimentar do ELSA-Brasil. Cad Saúde Pública 2013 Feb;29(2):379-389. [doi: 10.1590/s0102-311x2013000600024]

20. Marshall AL, Smith BJ, Bauman AE, Kaur S. Reliability and validity of a brief physical activity assessment for use by family doctors. Br J Sports Med 2005 May;39(5):294-297. [doi: 10.1136/bjsm.2004.013771] [Medline: 15849294]

21. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]

22. McKinney W. Data Structures for Statistical Computing in Python. 2010 Presented at: 9th Python in Science Conference (SciPy 2010); June 28 - July 3, 2010; Austin, TX. [doi: 10.25080/majora-92bf1922-00a]

23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Machine Learning Res 2011;12:2825-2830 [FREE Full text]

24. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]

25. Vigitel Brasil 2016: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico. Ministério da Saúde. 2016. URL: http://portalarquivos.saude.gov.br/images/pdf/2017/junho/07/vigitel_2016_jun17.pdf [accessed 2020-10-10]

26. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014 Jul 29;349:g4490. [doi: 10.1136/bmj.g4490] [Medline: 25073782]

27. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

28.   Lau Y, Chee DGH, Chow XP, Cheng LJ, Wong SN. Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials. Prev Med 2020 Mar;132:106001. [doi: 10.1016/j.ypmed.2020.106001] [Medline: 31991155]

29.   Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006 Aug;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]

30.   Teixeira PJ, Silva MN, Mata J, Palmeira AL, Markland D. Motivation, self-determination, and long-term weight control. Int J Behav Nutr Phys Act 2012;9:22 [FREE Full text] [doi: 10.1186/1479-5868-9-22] [Medline: 22385818]

31.   Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. Patient Prefer Adherence 2016;10:1547-1559 [FREE Full text] [doi: 10.2147/PPA.S103649] [Medline: 27574404]

32.   Mastellos N, Gunn LH, Felix LM, Car J, Majeed A. Transtheoretical model stages of change for dietary and physical exercise modification in weight loss management for overweight and obese adults. Cochrane Database Syst Rev 2014;2:CD008066. [doi: 10.1002/14651858.CD008066.pub3] [Medline: 24500864]

33.   Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 07;285(9):1172-1177. [doi: 10.1001/jama.285.9.1172] [Medline: 11231746]

34.   Alexander GL, McClure JB, Calvi JH, Divine GW, Stopponi MA, Rolnick SJ, et al. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. Am J Public Health 2010 Feb;100(2):319-326 [FREE Full text] [doi: 10.2105/AJPH.2008.154468] [Medline: 20019315]

Abbreviations

BCT: behavior change technique
POEmaS: Online Platform for Healthy Weight Loss

Original Paper

Barriers and Enablers to Adoption of Digital Health Interventions to Support the Implementation of Dietary Guidelines in Early Childhood Education and Care: Cross-Sectional Study

Alice Grady[1,2,3,4], PhD; Courtney Barnes[1,2,3,4], BNutrDiet; Luke Wolfenden[1,2,3,4], PhD; Christophe Lecathelinais[2], DESS; Sze Lin Yoong[1,2,3,4], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Population Health, Hunter New England Local Health District, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton, Australia

[4]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

Corresponding Author:
Alice Grady, PhD
School of Medicine and Public Health
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 0249246310
Email: alice.grady@health.nsw.gov.au

Abstract

Background: Few Australian childcare centers provide foods consistent with sector dietary guidelines. Digital health technologies are a promising medium to improve the implementation of evidence-based guidelines in the setting. Despite being widely accessible, the population-level impact of such technologies has been limited due to the lack of adoption by end users.

Objective: This study aimed to assess in a national sample of Australian childcare centers (1) intentions to adopt digital health interventions to support the implementation of dietary guidelines, (2) reported barriers and enablers to the adoption of digital health interventions in the setting, and (3) barriers and enablers associated with high intentions to adopt digital health interventions.

Methods: A cross-sectional telephone or online survey was undertaken with 407 childcare centers randomly sampled from a publicly available national register in 2018. Center intentions to adopt new digital health interventions to support dietary guideline implementation in the sector were assessed, in addition to perceived individual, organizational, and contextual factors that may influence adoption based on seven subdomains within the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) of health and care technologies framework. A multiple-variable linear model was used to identify factors associated with high intentions to adopt digital health interventions.

Results: Findings indicate that 58.9% (229/389) of childcare centers have high intentions to adopt a digital health intervention to support guideline implementation. The *changes needed in team interactions* subdomain scored lowest, which is indicative of a potential barrier (mean 3.52, SD 1.30), with *organization's capacity to innovate* scoring highest, which is indicative of a potential enabler (mean 5.25, SD 1.00). The two NASSS subdomains of *ease of the adoption decision* (*P*<.001) and *identifying work and individuals involved in implementation* (*P*=.001) were significantly associated with high intentions to adopt digital health interventions.

Conclusions: A substantial proportion of Australian childcare centers have high intentions to adopt new digital health interventions to support dietary guideline implementation. Given evidence of the effectiveness of digital health interventions, these findings suggest that such an intervention may make an important contribution to improving public health nutrition in early childhood.

XSL•FO
RenderX

Introduction

Poor diet is a modifiable risk factor and leading cause of burden of disease globally [1], with 11 million deaths and 255 million disability-adjusted life years in 2017 attributable to dietary risk factors [2]. Early childhood is a critical period to instill healthy eating habits to reduce diet-related burden of disease, as dietary behaviors developed in childhood track into adulthood [3]. Within Australia, population surveys indicate preschool-aged children are not consuming the recommended servings of fruits and vegetables, and consume more than recommended amounts of discretionary foods (ie, foods high in sodium, saturated fat, and added sugar) [4-6]. As a strategy to reduce the burden from poor diet, the implementation of dietary guidelines in the early childhood education and care (ECEC) setting is recommended [7,8]. Despite such recommendations, Australian childcare centers do not provide foods consistent with sector dietary guidelines [9-11]. For example, a 2017 audit of menus in 70 childcare centers across New South Wales (NSW) determined none of the menus were fully compliant with sector-specific dietary guidelines, particularly for vegetables [9].

Digital health interventions (eg, web-based programs, apps, etc) are advocated by the World Health Organization [12] and offer the opportunity to deliver support at scale and at low cost to improve the nutrition-related practices of food service organizations, such as childcare centers [13]. Evidence from randomized controlled trials [11,14-16] suggests that digital health interventions in education settings can improve the purchasing, provision, and consumption of healthier foods. Despite the clear potential of technology-based approaches, such interventions to improve health outcomes are often not adopted by end users, that is, the individual or organization for which the digital health intervention was developed (eg, ECEC centers, schools, and parents within educational settings). For example, it has been estimated that 80% of health technologies fail [17] due to uncertainty (ie, doubt about the technology's value or dependability), abandonment (ie, ceasing use of the technology), and lack of organizational willingness to adopt the technology [18] when disseminated in real-world contexts.

Broadly, systematic reviews, guidelines, and previous literature suggest that factors across a number of levels are important for the adoption and implementation of digital health interventions. These include factors related to the individual user (eg, knowledge, skills, beliefs, and attitudes) [19-21], the organization (eg, compatibility or fit with the organization, access to appropriate infrastructure and equipment, and leadership engagement) [12,19-21], the wider setting (eg, external policies and incentives) [12,19,21], the process of implementation (eg, lack of considered planning, engagement, and evaluation) [21], and the technology-based intervention itself (eg, complexity, costs, adaptability, and ability of the intervention to meet user needs) [12,19-21].

Within ECEC settings, a 2015 systematic review examining the barriers to integration of information technology more broadly, including computers, tablets, and touchscreen whiteboards, identified a scarcity of empirical studies examining barriers and enablers within the setting, none of which focused on improving

guideline implementation or child health outcomes [19]. The lack of research examining the factors that may enable or impede the adoption of digital health interventions to improve dietary guideline implementation is problematic, as such evidence is necessary to inform future strategies to maximize the adoption and, therefore, impact of evidence-based technologies in the setting.

As such, by employing the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) of health and care technologies framework [17], this study aimed to describe the following in a randomly selected national sample of Australian childcare centers: (1) intentions to adopt digital health interventions to support childcare center implementation of dietary guidelines, (2) reported individual, organizational, and contextual barriers and enablers to the adoption of digital health interventions in the setting, and (3) barriers and enablers associated with high intentions to adopt digital health interventions.

Methods

Study Design, Ethics Approval, and Consent to Participate

This study employed a cross-sectional design. Ethical approval was obtained by the Human Research Ethics Committees of Hunter New England (16/02/17/4.05) and the University of Newcastle (H-2016-0111). All subjects in this research study provided consent to participate.

Sample

The Australian Children's Education and Care Quality Authority's (ACECQA) national register [22] was used to obtain a sampling frame of potentially eligible center-based childcare centers, including long day cares (ie, centers that provide care for children aged 0-6 years for >8 hours per day) and preschools (ie, centers that provide care for children aged 3-6 years for 6-8 hours per day) [5], from each state within Australia (N=10,631). A sample of 1500 childcare centers (14.11%) were randomly selected from the sampling frame of potentially eligible centers, stratified by state and center area socioeconomic classification by an independent statistician.

Childcare center eligibility was assessed via online or telephone survey items. Centers were deemed ineligible if they did not provide meals to children or make menu planning decisions onsite, as this survey was assessing technology to support nutrition guideline implementation on menus; had staff with insufficient English to complete the survey; were a Department of Education and Communities center, as ethical approval was not obtained from the relevant government department; were located in the Hunter New England region of NSW or were select centers across NSW, due to concurrent nutrition and physical activity research trials being undertaken by the research team; were identified as out-of-school hours, vacation care, or family day care; or catered solely to children with special needs.

Recruitment and Procedures

An email with an information statement and link to an online survey was sent to the nominated supervisor (ie, the center

manager) of all sampled childcare centers (N=1500) inviting them to assess eligibility and participate in the study. Nominated supervisors were able to select an alternate staff member (eg, center director) to complete the survey on their behalf. Centers that did not complete the survey within 4 weeks were sent a reminder email to participate (1466/1500, 97.73%), followed by a phone call from a member of the research team (1455/1500, 97.00%) to assess eligibility and gain verbal consent to complete the telephone version of the survey. A final reminder email was sent to centers that indicated a preference to complete the online version of the survey (846/1500, 56.40%) and those who were noncontactable via phone. Centers that were yet to complete the survey following the final reminder email received a final telephone call to gain consent and complete a telephone version of the survey (744/1500, 49.60%). Centers were not offered any incentives to complete the survey. Data to assess study outcomes were collected between January and August 2018.

Data Collection and Measures

Center and Responder Characteristics

Childcare centers were asked to report on the type of center (ie, preschool or long day care), number of full-time equivalent staff members, center opening and closing hours, number of children enrolled, and the number of children enrolled identifying as of Aboriginal and/or Torres Strait Islander background. Childcare center staff completing the survey were asked to report their main role at the center and the total number of years working in the childcare setting. Survey items assessing center characteristics were sourced from previous Australian childcare center surveys conducted by the research team [11,23,24].

Center geographical information, including state and postcode, were obtained via the ACECQA national register to determine location and the center area socioeconomic classification.

Intentions to Adopt Digital Health Interventions

To aid comprehension and standardization of digital health interventions and their capabilities, participants were first given a brief example of the potential modality (eg, web-based or online) and key features (eg, feedback and tips) that could be provided within a digital health intervention to support guideline implementation in the setting. Three survey items derived from the Technology Acceptance Model [25] were then used to assess childcare centers' intentions to adopt digital health interventions in the setting. The Technology Acceptance Model is an information systems theory that models how end users come to accept and use a new technology [25]. The Technology Acceptance Model has been shown to have high internal consistency (Cronbach α>.80) [26]. Respondents were asked to rate on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), how much they agreed with the following statements, assuming they had access to a digital health intervention to support the implementation of dietary

guidelines in their center: "I intend to use it," "I predict that I would use it," and "I would plan to use it." These items have been used in previous research by the team in the ECEC setting [27].

Barriers and Enablers to Adoption of Digital Health Interventions

A purpose-built measure based on the NASSS framework by Greenhalgh [18] was used to assess individual, organizational, and contextual factors that may influence adoption of digital health technologies to improve the implementation of dietary guidelines in the childcare setting. The NASSS framework is an evidence-based, theory-informed, and pragmatic framework designed to help predict and evaluate the success of a technology-supported health care program [18]. The NASSS consists of seven domains: the illness or condition, the technology, the value proposition, the individuals intended to adopt the technology, the organizations, the wider system, and how all these domains interact over time [28]. The NASSS framework can be used to generate insight into the multiple influences on the success or failure of a complex technology-based intervention; to identify simple, complicated, and complex components of the intervention; and to consider how individuals and organizations may be supported to handle complex components of the intervention [18].

An expert advisory group, including health promotion practitioners, implementation scientists, and dietitians, was involved in the development of the measure. Based on expert advisory group consensus, only three of the seven NASSS domains were deemed relevant to the end users for the scale of dissemination of digital health interventions under examination and were, therefore, assessed. At the time of survey development, no validated measure for the NASSS framework existed. As such, a search was conducted for validated measures that had corresponding domains to the NASSS framework. Where possible, such validated measures were employed and adapted to fit the ECEC context, including the e-Health Readiness Measure [29], which was utilized for two of the subdomains: the organization's capacity to innovate and readiness of the organization for technology-supported change. The e-Health Readiness Measure [30] has been shown to have high internal consistency (Cronbach α>.80). Items for the remaining five subdomains were developed by the advisory group and pilot-tested among a group of health promotion practitioners and trained telephone interviewers for comprehension and face validity. The final measure consisted of 24 items, 10 of which were adapted items from the e-Health Readiness Measure, across three domains and seven subdomains of the NASSS framework, rated on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). Table 1 lists the domains, subdomains, number of survey items, and an example survey item relevant to the setting.

Table 1. Nonadoption, abandonment, scale-up, spread, and sustainability framework application to the early childhood education and care setting.

Domain and subdomain (No. of items)	Example survey item
The adopter system	
Changes in staff roles, practices, and identities (3 items)	Using an online program is consistent with the usual practices of my cook and menu planner.
The organization	
Organization's capacity to innovate (6 items)	Overall, I think our service has a champion or leader for using new technology.
Readiness of the organization for technology-supported change (4 items)	Overall, I think our service has access to experts in use of new technology.
Ease of the adoption and funding decision (1 item)	It would be easy to adopt new technology to support menu planning in my service.
Changes needed in team interactions and routines (2 items)	My service would need to change the way it currently plans menus if we decided to adopt new technology.
Identifying work and individuals involved in implementation (2 items)	We already have the existing personnel available to support the adoption of new technology.
The wider context	
Political, economic, regulatory, professional (eg, medicolegal), and sociocultural context for program rollout (6 items)	I would be more likely to adopt new technology in my service if it was promoted by relevant government agencies (ie, Department of Education or Department of Health).

Analysis

Overview

All analyses were performed in SAS, version 9.3 (SAS Institute) [31]. Descriptive statistics including means, frequencies, and proportions were used to describe center demographic characteristics and survey responses. Childcare center postcodes ranked in the top 50% of NSW, according to the Socio-Economic Indexes for Areas, were classified as *higher socioeconomic status* [32]. A chi-square analysis (ie, test of independence) was used to compare center area socioeconomic classification among consenters and nonconsenters.

Intentions to Adopt Digital Health Interventions

An intention-to-adopt score for each responder was calculated by averaging scores for the three intention items. This score was also used to dichotomize responders into having low intentions to adopt (score <6) or high intentions to adopt (score ≥6). This cut point corresponds to those who agree or strongly agree with each item. Such an approach has been used previously within ECEC centers [27].

Barriers and Enablers to Adoption of Digital Health Interventions

Similar to previous studies assessing barriers and enablers using theoretical frameworks [9,33-35], average scores for each NASSS construct were calculated by summing all scores for all items within the subdomain, ranging from 1 (strongly disagree) to 7 (strongly agree), and dividing by the total number of responses within the domain. Six survey items were negatively worded and were, therefore, reverse scored. Mean values were used to describe the domains as potential barriers and enablers [33]. A lower mean (≤4) suggested that the particular domain may be a barrier, and a higher mean (>4) suggested a perceived enabler to adoption of digital health

interventions. In consultation with the expert advisory group (ie, health promotion practitioners, implementation scientists, and dieticians), this cutoff was employed as a pragmatic approach to categorizing mean scores (ie, ≤4 [responses strongly disagree to neither agree nor disagree] and >4 [responses slightly agree to strongly agree]) and was chosen to limit reporting of any potential social desirability bias in the identification of enablers.

Barriers and Enablers Associated With Intentions to Adopt Digital Health Interventions

All seven NASSS subdomains were entered as independent variables into a multiple-variable logistic regression model, to assess which NASSS constructs were significantly associated with high intentions to adopt digital health interventions (ie, dependent variable) after adjusting for each other. The significance value was set at .05.

Results

Characteristics of Participants

Of the 1500 centers invited to participate in the study, 72 (4.80%) were noncontactable, 53 (3.53%) were contacted but did not respond, and 378 (25.20%) declined to participate prior to eligibility being assessed. A total of 997 out of 1500 (66.47%) centers consented to the study and were assessed for eligibility, with 590 of these 997 (59.2%) centers deemed ineligible, most commonly due to the center not providing meals and/or snacks to children and being a Department of Education and Community center. This resulted in a total of 407 centers taking part in the survey. There were no statistically significant differences in center socioeconomic area between consenters and nonconsenters.

The large majority of participating centers were long day care centers (391/407, 96.1%) (see Table 2). The majority of

responders held the position of nominated supervisor (183/399, 45.9%) or director (179/399, 44.9%), with more than 10 years'

experience working in the childcare setting (278/397, 70.0%).

Table 2. Childcare center and responder characteristics.

Characteristics	Value, n (%) or mean (SD)
Center characteristics (N=407)	
Type of center, n (%)	
Preschool	16 (3.9)
Long day care center	391 (96.1)
Number of children enrolled (n=406), mean (SD)	96.33 (56.79)
Number of full-time equivalent primary contact teaching staff (n=404), mean (SD)	12.78 (7.93)
Number of children of Aboriginal and/or Torres Strait Islander background enrolled at center (n=406), n (%)	214 (52.7)
Center area socioeconomic status , n (%)	
High	231 (56.8)
Low	176 (43.2)
Center geographic location , n (%)	
Urban (major cities)	307 (75.4)
Rural (inner regional, outer regional, or remote)	100 (24.6)
Center state, n (%)	
New South Wales	165 (40.5)
Victoria	94 (23.1)
Queensland	62 (15.2)
Australian Capital Territory	7 (1.7)
Tasmania	7 (1.7)
Western Australia	40 (9.8)
South Australia	25 (6.1)
Northern Territory	7 (1.7)
Responder characteristics (n=399), n (%)	
Role at the center	
Nominated supervisor	183 (45.9)
Director	179 (44.9)
Cook	12 (3.0)
Other	28 (7.0)
Number of years working in the childcare setting (n=397), n (%)	
≤5	36 (9.1)
6-10	83 (20.9)
>10	278 (70.0)

Intentions to Adopt Digital Health Interventions

The mean intention score was 5.52 (SD 1.07), with a median of 6.00 (IQR 5.00-6.00). Of 389 responders, 229 (58.9%) centers had high intentions to adopt digital health interventions to support the implementation of dietary guidelines.

Reported Barriers and Enablers to Adoption of Digital Health Interventions

A mean score of 4 or lower (ie, barriers) was found for four of the seven NASSS domains (see Table 3). For three of the seven NASSS constructs—organization's capacity to innovate, ease of the adoption and funding decision, and political context—responders had mean scores of more than 4 (ie, enablers).

XSL•FO
RenderX

Table 3. Mean and median scores for the nonadoption, abandonment, scale-up, spread, and sustainability subdomain barriers and enablers, as reported by responders.

Barrier or enabler	Score[a]	
	Mean (SD)[b]	Median (IQR)
The adopter system (n=390)		
Changes in staff roles, practices, and identities	4.32 (1.25)	4.33 (3.33-5.00)
The organization		
Organization's capacity to innovate (n=382)	5.25 (1.00)	5.50 (4.67-6.00)
Readiness of the organization for technology-supported change (n=386)	4.88 (1.03)	5.00 (4.25-5.75)
Ease of the adoption and funding decision (n=387)	5.22 (1.31)	6.00 (4.00-6.00)
Changes needed in team interactions and routines (n=389)	3.52 (1.30)	3.50 (2.50-4.00)
Identifying work and individuals involved in implementation (n=389)	4.35 (1.19)	4.00 (4.00-5.00)
The wider context (n=389)		
Political, economic, regulatory, professional (eg, medicolegal), and sociocultural context for program rollout	5.07 (1.08)	5.33 (4.50-6.00)

[a]Constructs are reported on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree).

[b]A mean of ≤4 suggests that the particular domain may be a barrier; a mean of >4 suggests the domain may be an enabler.

Barriers and Enablers Associated With Adoption of Digital Health Interventions

Multiple-variable logistic regression analyses revealed a significant association between two of the NASSS subdomains and high intentions to adopt digital health interventions (see Table 4). For every 1-point increase in the *ease of the adoption* *and funding decision* subdomain, centers were 1.75 times more likely to have high intentions of adopting digital health interventions (95% CI 1.40-2.18; *P*<.001). For every 1-point increase in the *identifying work and individuals involved in implementation* subdomain, centers had 1.46 times the odds of having high intentions to adopt digital health interventions (95% CI 1.61-1.84; *P*=.001).

Table 4. Nonadoption, abandonment, scale-up, spread, and sustainability subdomains associated with high intentions to adopt digital health interventions in early childhood education and care centers.

Barrier or enabler	Odds ratio	95% CI	*P* value
The adopter system			
Changes in staff roles, practices, and identities	0.88	0.71-1.10	.27
The organization			
Organization's capacity to innovate	1.26	0.91-1.75	.17
Readiness of the organization for technology-supported change	1.15	0.83-1.59	.41
Ease of the adoption and funding decision	1.75	1.40-2.18	<.001
Changes needed in team interactions and routines	0.92	0.75-1.13	.42
Identifying work and individuals involved in implementation	1.46	1.16-1.84	.001
The wider context			
Political, economic, regulatory, professional (eg, medicolegal), and sociocultural context for program rollout	1.03	0.82-1.29	.81

Discussion

Principal Findings

This novel study applied a technology-specific framework to conduct a theoretical assessment of childcare center barriers and enablers to the adoption of digital health interventions to improve dietary guideline implementation, nationally. Application of the NASSS framework resulted in the identification of a number of reported barriers and enablers.

The main barrier identified was *changes needed in team interactions and routines*, with the main enablers identified as being *ease of the adoption decision*, *identifying work and individuals involved in implementation*, and *organization's capacity to innovate*. Centers that reported higher scores in the *ease of the adoption decision* and the *identifying work and individuals involved in implementation* subdomains were significantly more likely to have high intentions of adopting digital health interventions.

XSL•FO
RenderX

The study found that over half (229/389, 58.9%) of responders had high intentions to adopt digital health interventions in the setting. Few studies of technology-based health interventions within the ECEC setting report adoption rates, with variable findings. A 2015 cross-sectional study assessing intentions to adopt a web-based program to support healthy eating and physical activity policies and practices in the ECEC setting reported that 72% of respondents had high intentions to adopt such a program [27]. In our earlier study assessing the impact of implementation support on actual adoption of a web-based menu planning program, 58% of the control group, who did not receive support, had adopted the program [36]. Combined, these findings are indicative of the relatively high intentions to adopt digital health technologies in the ECEC setting.

When examining the potential barriers and enablers to adoption of digital health interventions, scores of 4 or higher were found for only three of the subdomains assessed (ie, enablers), two of which fall within the organizational construct of the NASSS framework. The highest levels of agreement were found for the *organization's capacity to innovate* (mean 5.25, SD 1.00), the *ease of the adoption decision* (mean 5.22, SD 1.31), and *political context* (mean 5.07, SD 1.08). This suggests these subdomains may be potential enablers of the adoption of digital health interventions for end users. Responders reported the lowest level of agreement for *changes needed in team interactions* (mean 3.52, SD 1.30) within the organizational construct, which suggests this subdomain may be a potential barrier to adoption. Such findings suggest that in order to facilitate the adoption of new technology, strategies that generate a high level of organizational support (eg, informing opinion leaders, involving executive boards, and mandating change) and those that overcome any operational challenges and changes in practice (eg, educational outreach visits, changing equipment, and local technical assistance) should be considered [37]. This finding is consistent with previous research demonstrating that implementation support strategies, including face-to-face training, ongoing telephone support, and provision of resources and infrastructure, in addition to obtaining managerial support, improved the adoption of a web-based program in the setting [36].

Study results revealed a discrepancy in the reported barriers and enablers to adoption of digital health interventions and the factors associated with adoption. Multiple-variable logistic regression analyses determined that the *ease of the adoption decision* and the *identifying work and individuals involved in implementation* subdomains were the only factors to have a statistically significant association with high intentions to adopt. Responders scoring higher, that is, those with greater agreement, on these two factors were 1.75 and 1.46 times more likely to report high intentions to adopt digital health interventions, respectively. Although previous studies have not specifically assessed such theoretical constructs in this setting, incongruity in the perceived versus the actual experiences of barriers to the adoption of technology-based interventions [19] and evidence-based guidelines [9] within the ECEC setting has been reported previously. There are opportunities to target this identified incongruence. In-depth examination of the factors by way of supplementation with qualitative methods among all

intended end users is warranted. This may provide greater insights into the complexities to adoption of technology-based health interventions and the interaction between each domain of the NASSS.

While recent studies have employed the NASSS framework retrospectively to categorize various constructs [38,39], this study is novel in its prospective application of the NASSS as a measure to conduct a theory-based assessment. Future research could further examine use of the NASSS as a tool to identify barriers and enablers to the adoption of digital health interventions to inform intervention development and evaluation. In addition, embedding measures of the NASSS into the evaluation of dissemination interventions to improve adoption of digital health interventions would allow for an examination of mechanisms and provide a better understanding of how individual, organizational, and contextual factors impact adoption.

Limitations

The intention to adopt digital health interventions, rather than actual adoption, was assessed. While there is evidence of a relationship between intentions and actual adoption [40], and while our findings align with prior research in the setting [36], rates of actual adoption may differ to those reported. While drawing on validated measures used in other settings, this study employed a nonvalidated self-reported measure to assess barriers and enablers to the adoption of digital health interventions, which may be subject to social desirability bias [41]. Three of the subdomains—*changes needed in team interaction, identifying work and individuals involved in implementation,* and *ease of funding decision*—contained less than three items and, as such, should be interpreted with caution. This study also did not assess additional contextual factors that are theorized to influence adoption according to the NASSS framework, such as the condition, the technology, the value proposition, and embedding and adapting over time [18]. Future studies should consider undertaking an assessment of such factors to assist in providing a more comprehensive understanding of the broader factors that may impact adoption of digital health technologies in the childcare setting. Finally, as Department of Education and Community centers were not eligible to participate, study findings may not be representative of these centers. However, as the geographic distribution of participating centers is similar to that of the sampling frame—all center-based childcare within the ACECQA national register (differences between the responders in each state vs state population ranged from 0.32% to 7.05%)—the sample may be considered nationally representative.

Conclusions

This study provides novel insights into the perceived and actual factors that may facilitate or impede the adoption of digital health interventions at scale from the perspective of end users. A substantial proportion of Australian childcare centers reported high intentions to adopt digital health interventions. Given evidence of the effectiveness of such technologies, these interventions have the potential to make an important contribution to improving public health nutrition in early childhood. Nonetheless, future efforts to disseminate digital

health prevention programs at scale should consider targeting factors within the *ease of the adoption decision* and *identifying* *work and individuals involved in implementation* subdomains in order for adoption to be ubiquitous in the setting.

Authors' Contributions

All authors contributed to conception or design of the work, data acquisition, and analysis or interpretation of data, and took part in revising the manuscript. All authors give their final approval of this version to be published and agree to be accountable for all aspects of the work. AG, SY, and LW conceived the study and secured funding. AG and SY designed the evaluation procedures. AG and CB lead the acquisition of data. CL conducted the data analysis. AG led the drafting of the manuscript.

Conflicts of Interest

None declared.

References

1. Lim S, Vos T, Flaxman A, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2224-2260. [doi: 10.3410/f.719894684.793533485]

2. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]

3. Craigie A, Lake A, Kelly S, Adamson A, Mathers J. Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]

4. Australian Bureau of Statistics. Australian Health Survey: Consumption of Food Groups From the Australian Dietary Guidelines. Australia 2011-12. Canberra, Australia: Commonwealth of Australia; 2016. URL: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/77E3E65E4D967249CA2581F40016ABD6/$File/4364.0.55.012%20-%20consumption%20of%20food%20groups%20from%20australian%20dietary%20guidelines.pdf [accessed 2020-11-13]

5. Baxter J. Child care and early childhood education in Australia. Australian Institute of Family Studies. 2015 May. URL: https://aifs.gov.au/publications/child-care-and-early-childhood-education-australia [accessed 2018-12-01]

6. Centre for Epidemiology and Evidence. Fruit and vegetables: Recommended daily consumption by children. HealthStats NSW. Sydney, Australia: NSW Ministry of Health; 2020 May. URL: http://www.healthstats.nsw.gov.au/Indicator/beh_vegkid_age/beh_vegkid_age_snap [accessed 2018-12-01]

7. Institute of Medicine (IOM), Committee on Obesity Prevention Policies for Young Children. In: Birch LL, Parker L, Burns A, editors. Early Childhood Obesity Prevention Policies. Washington, DC: The National Academies Press; 2011.

8. Global Nutrition Targets 2025: Policy brief series. World Health Organization. Geneva, Switzerland: World Health Organization; 2014. URL: https://www.who.int/nutrition/publications/globaltargets2025_policybrief_overview/en/ [accessed 2020-11-13]

9. Grady A, Seward K, Finch M, Fielding A, Stacey F, Jones J, et al. Barriers and enablers to implementation of dietary guidelines in early childhood education centers in Australia: Application of the theoretical domains framework. J Nutr Educ Behav 2018 Mar;50(3):229-237.e1. [doi: 10.1016/j.jneb.2017.09.023] [Medline: 29170057]

10. Yoong S, Skelton E, Jones J, Wolfenden L. Do childcare services provide foods in line with the 2013 Australian dietary guidelines? A cross-sectional study. Aust N Z J Public Health 2014 Dec;38(6):595-596. [doi: 10.1111/1753-6405.12312] [Medline: 25440467]

XSL·FO

RenderX

11. Grady A, Wolfenden L, Wiggers J, Rissel C, Finch M, Flood V, et al. Effectiveness of a web-based menu-planning intervention to improve childcare service compliance with dietary guidelines: Randomized controlled trial. J Med Internet Res 2020 Feb 04;22(2):e13401 [FREE Full text] [doi: 10.2196/13401] [Medline: 32014843]

12. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-11-13]

13. Vandelanotte C, Müller AM, Short C, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]

14. Sutherland R, Nathan N, Brown A, Yoong S, Finch M, Lecathelinais C, et al. A randomized controlled trial to assess the potential efficacy, feasibility and acceptability of an m-health intervention targeting parents of school aged children to improve the nutritional quality of foods packed in the lunchbox 'SWAP IT'. Int J Behav Nutr Phys Act 2019 Jul 02;16(1):54 [FREE Full text] [doi: 10.1186/s12966-019-0812-7] [Medline: 31266506]

15. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, et al. Cluster randomized controlled trial of a consumer behavior intervention to improve healthy food purchases from online canteens. Am J Clin Nutr 2017 Nov;106(5):1311-1320. [doi: 10.3945/ajcn.117.158329] [Medline: 28971849]

16. Yoong S, Grady A, Wiggers J, Stacey FG, Rissel C, Flood V, et al. Child-level evaluation of a web-based intervention to improve dietary guideline implementation in childcare centers: A cluster-randomized controlled trial. Am J Clin Nutr 2020 Apr 01;111(4):854-863 [FREE Full text] [doi: 10.1093/ajcn/nqaa025] [Medline: 32091593]

17. Greenhalgh T. Why do so many technology projects in healthcare fail? A new framework for studying the non-adoption, abandonment and failure of scale-up, spread and sustainability (NASSS) of health and care technologies. Oxford Talks. 2018. URL: https://new.talks.ox.ac.uk/talks/id/4e41675d-317c-467f-8c70-96918ca171eb [accessed 2018-12-01]

18. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]

19. Plumb M, Kautz K. Barriers to the integration of information technology within early childhood education and care organisations: A review of the literature. In: Proceedings of the Australasian Conference on Information Systems. 2016 Presented at: Australasian Conference on Information Systems; November 30-December 4, 2015; Adelaide, Australia.

20. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson SA. Computerized clinical decision support for prescribing: Provision does not guarantee uptake. J Am Med Inform Assoc 2010;17(1):25-33 [FREE Full text] [doi: 10.1197/jamia.M3170] [Medline: 20064798]

21. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: A systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]

22. National registers. Australian Children's Education and Care Quality Authority. URL: https://www.acecqa.gov.au/resources/national-registers [accessed 2020-11-13]

23. Finch M, Seward K, Wedesweiler T, Stacey F, Grady A, Jones J, et al. Challenges of increasing childcare center compliance with nutrition guidelines: A randomized controlled trial of an intervention providing training, written menu feedback, and printed resources. Am J Health Promot 2019 Mar;33(3):399-411. [doi: 10.1177/0890117118786859] [Medline: 30004247]

24. Yoong SL, Finch M, Nathan N, Wiggers J, Lecathelinais C, Jones J, et al. A longitudinal study assessing childcare services' adoption of obesity prevention policies and practices. J Paediatr Child Health 2016 Jul;52(7):765-770. [doi: 10.1111/jpc.13252] [Medline: 27244222]

25. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]

26. King WR, He J. A meta-analysis of the technology acceptance model. Inf Manage 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]

27. Yoong SL, Williams CM, Finch M, Wyse R, Jones J, Freund M, et al. Childcare service centers' preferences and intentions to use a web-based program to implement healthy eating and physical activity policies and practices: A cross-sectional study. J Med Internet Res 2015 Apr 30;17(5):e108 [FREE Full text] [doi: 10.2196/jmir.3639] [Medline: 25931430]

28. Greenhalgh T, Abimbola S. The NASSS framework - A synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]

29. Touré M, Poissant L, Swaine BR. Assessment of organizational readiness for e-health in a rehabilitation centre. Disabil Rehabil 2012;34(2):167-173. [doi: 10.3109/09638288.2011.591885] [Medline: 21936712]

30. Poissant L, Curran J. The development of a questionnaire to assess organizational readiness to adopt e-health technologies. In: Proceedings of the Annual Canadian Association for Health Services and Policy Research (CAHSPR) Conference. Ottawa, ON: Canadian Association for Health Services and Policy Research; 2007 Presented at: Annual Canadian Association for Health Services and Policy Research (CAHSPR) Conference; June 2007; Toronto, ON p. 33.

31. SAS 9.3 software. SAS Support. Cary, NC: SAS Institute Inc URL: https://support.sas.com/software/93/#s1=1 [accessed 2020-12-13]

32. Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA) 2016. Canberra, Australia: Commonwealth of Australia; 2018. URL: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf [accessed 2020-11-13]

33. Beenstock J, Sniehotta FF, White M, Bell R, Milne EM, Araujo-Soares V. What helps and hinders midwives in engaging with pregnant women about stopping smoking? A cross-sectional survey of perceived implementation difficulties among midwives in the North East of England. Implement Sci 2012 Apr 24;7:36 [FREE Full text] [doi: 10.1186/1748-5908-7-36] [Medline: 22531641]

34. Taylor N, Lawton R, Conner M. Development and initial validation of the determinants of physical activity questionnaire. Int J Behav Nutr Phys Act 2013 Jun 11;10:74 [FREE Full text] [doi: 10.1186/1479-5868-10-74] [Medline: 23758912]

35. Taylor N, Parveen S, Robins V, Slater B, Lawton R. Development and initial validation of the Influences on Patient Safety Behaviours Questionnaire. Implement Sci 2013 Jul 29;8:81 [FREE Full text] [doi: 10.1186/1748-5908-8-81] [Medline: 23895628]

36. Grady A, Wolfenden L, Rissel C, Green S, Reilly K, Yoong SL. Effectiveness of a dissemination strategy on the uptake of an online menu planning program: A controlled trial. Health Promot J Austr 2019 Dec;30 Suppl 1:20-25. [doi: 10.1002/hpja.220] [Medline: 30417473]

37. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]

38. Dijkstra A, Heida A, van Rheenen PF. Exploring the challenges of implementing a web-based telemonitoring strategy for teenagers with inflammatory bowel disease: Empirical case study. J Med Internet Res 2019 Mar 29;21(3):e11761 [FREE Full text] [doi: 10.2196/11761] [Medline: 30924785]

39. Edridge C, Deighton J, Wolpert M, Edbrooke-Childs J. The implementation of an mHealth intervention (ReZone) for the self-management of overwhelming feelings among young people. JMIR Form Res 2019 May 02;3(2):e11958 [FREE Full text] [doi: 10.2196/11958] [Medline: 31045499]

40. Turner M, Kitchenham B, Brereton P, Charters S, Budgen D. Does the technology acceptance model predict actual use? A systematic literature review. Inf Softw Technol 2010 May;52(5):463-479. [doi: 10.1016/j.infsof.2009.11.005]

41. Grimm P. Social desirability bias. In: Kamakura WA, editor. Wiley International Encyclopedia of Marketing. Volume 2: Marketing Research. Chichester, UK: John Wiley & Sons; 2010:258-259.

Abbreviations

ACECQA: Australian Children's Education and Care Quality Authority
CCNSW: Cancer Council NSW
ECEC: early childhood education and care
HNEPH: Hunter New England Population Health
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
NHMRC: National Health and Medical Research Council
NSW: New South Wales
TAPPC: The Australian Prevention Partnership Centre

XSL•FO
RenderX

Original Paper

Importance of Multiple Reinforcing Comments and Areas for Change in Optimizing Dietary and Exercise Self-Monitoring Feedback in Behavioral Weight Loss Programs: Factorial Design

Rebecca Krukowski[1], PhD; Hyeonju Kim[1], PhD; Melissa Stansbury[2], MS; Qian Li[3], MSc; Saunak Sen[1], PhD; Gregory Farage[1], PhD; Delia West[2], PhD

[1]Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[3]Department of Psychology, University of Memphis, Memphis, Memphis, TN, United States

Corresponding Author:
Rebecca Krukowski, PhD
Department of Preventive Medicine
College of Medicine
University of Tennessee Health Science Center
66 N. Pauline
Memphis, TN, 38163
United States
Phone: 1 9014482426
Email: rkrukows@uthsc.edu

Abstract

Background: Individualized dietary and physical activity self-monitoring feedback is a core element of behavioral weight loss interventions and is associated with clinically significant weight loss. To our knowledge, no studies have evaluated individuals' perspectives on the composition of feedback messages or the effect of feedback composition on the motivation to self-monitor.

Objective: This study aims to assess the perceptions of feedback emails as a function of the number of comments that reinforce healthy behavior and the number of areas for change (ie, behavioral changes that the individual might make to have an impact on weight) identified.

Methods: Emailed feedback followed a factorial design with 2 factors (ie, reinforcing comments and areas for change), each with 3 levels (ie, 1, 4, or 8 comments). A total of 250 adults with overweight or obesity who were interested in weight loss were recruited from the Qualtrics research panel. Participants read 9 emails presented in a random order. For each email, respondents answered 8 questions about the likelihood to self-monitor in the future, motivation for behavioral change, and perceptions of the counselor and the email. A mixed effects ordinal logistic model was used to compute conditional odds ratios and predictive margins (ie, average predicted probability) on a 5-point Likert response scale to investigate the optimal combination level of the 2 factors.

Results: Emails with more reinforcing comments or areas for change were better received, with small incremental benefits for 8 reinforcing comments or areas for change versus 4 reinforcing comments or areas for change. Interactions indicated that the best combination for 3 of 8 outcomes assessed (ie, motivation to make behavioral changes, counselor's concern for their welfare, and the perception that the counselor likes them) was the email with 8 reinforcing comments and 4 areas for change. Emails with 4 reinforcing comments and 4 areas for change resulted in the highest average probability of individuals who reported being very likely to self-monitor in the future.

Conclusions: The study findings suggest how feedback might be optimized for efficacy. Future studies should explore whether the composition of feedback email affects actual self-monitoring performance.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Individualized dietary and physical activity self-monitoring feedback on the frequency of monitoring, reinforcing comments about positive behaviors, and specific recommendations for behavioral change based on self-monitoring data is a core element of behavioral weight loss interventions [1-3]. Furthermore, Tate et al [4] found that human-generated emails produce significantly greater weight loss over 6 months than computer-automated feedback, which may reflect the benefit of tailored messages in general [5]. Despite the clear benefits of self-monitoring feedback in behavioral weight control programs, surprisingly little empirical information exists to guide the crafting of feedback messages. Some have suggested a formula for feedback, which *sandwiches* comments that reinforce positive health behaviors with suggested areas of change [6]. Others have argued that providing a *menu* of areas for change, which encompasses multiple dietary and physical activity behaviors to give individuals the option to choose what they would like to adopt, if any, is preferable to a prescriptive approach that targets one area to change [7].

To our knowledge, no studies have evaluated how individuals who attempt weight control perceive the effect of feedback messages on their self-monitoring or whether feedback composition affects the likelihood to self-monitor in the future or strengthens the relationship between the participant and the interventionist. As counselor time is one of the most expensive aspects of a behavioral weight management intervention [8,9], it is critical to examine whether longer, personalized feedback (requiring greater time to craft) would prompt continued self-monitoring and begin to identify the composition of feedback that would be expected to promote subsequent self-monitoring. Optimizing self-monitoring feedback may strengthen the working relationship between counselors and participants and, ultimately, increase weight loss.

From a theoretical perspective, feedback that reinforces positive health behaviors would be expected to build self-efficacy for behavioral change and provide positive outcome expectations for self-monitoring [10,11]. Given the strong positive association between greater self-monitoring and greater weight loss [12], outcome expectations that support self-monitoring and enhance self-efficacy for weight loss–promoting behaviors should promote better weight loss. Identifying multiple suggested areas for behavioral change supports autonomy because it gives the individual an option to select from these areas [13], and increased autonomy is also associated with better weight loss outcomes [14,15].

Objectives

Therefore, the purpose of this study is to assess how individuals perceive self-monitoring feedback emails (ie, likelihood to self-monitor in the future, motivation to make behavioral changes, and perceptions of the counselor and emails) depending on the number of comments on positive health behaviors and the number of identified areas for change. This study also examines whether emails would be better received if they had a greater number of reinforcing comments than areas for change.

We predicted that emails with the highest number of reinforcing comments (ie, 8) would be perceived as optimal for encouraging treatment engagement, but a moderate number of areas for change (ie, 4 areas for change) would be preferred.

Methods

Design

A total of 250 participants were recruited from the Qualtrics research panel, a web-based survey platform available to researchers to facilitate participant recruitment and web-based data collection, from April to May 2019 [16]. Qualtrics partners with more than 20 web-based panel providers; these panel partners randomly select respondents from their cadre of individuals and send them invitations to participate in surveys. Individuals completed a screening questionnaire to identify those who were eligible (ie, aged ≤18 years, had a BMI ≥25 kg/m^2, and a desire to lose weight). To ensure a demographically balanced survey panel, the number of participants within the gender and race strata (ie, individuals who identified as White and those who identified as another racial group) was capped. Qualtrics seeks to identify fraudulent respondents when taking surveys by methods such as monitoring respondents' survey speed and internet protocol addresses. Respondents were given point-based incentives by the panel provider that could be redeemed in various ways (eg, airline miles, credit for web-based games, and gift cards) for completing the survey. This study was approved by the institutional review board at the University of Tennessee Health Science Center, and participants provided informed consent.

Respondents were asked to imagine that they were in a weight loss program in which they recorded their diet and exercise daily, and they were told that they received feedback about their self-recorded diet and exercise from a counselor weekly via the following emails. Next, respondents read 9 emails in a computer-randomized order. The feedback presented in the emails followed a full factorial design with 2 factors (ie, reinforcing comments and areas for change), each with 3 levels (ie, 1 comment, 4 comments, and 8 comments).

After reading each email, respondents answered 8 questions about their likelihood to self-monitor in the future, motivation to make behavioral changes, and their perceptions of the counselor and the emails. Respondents spent an average of 27 (SE 3) min to complete the survey.

Emails

Archived email feedback on diet and exercise self-monitoring that had been previously written for participants in the iREACH3 behavioral weight control study [17] was used in this study. A pool of emails was classified into points that noted positive behaviors and those that suggested an area of change to consider. A total of 9 emails were crafted to meet the requirements of the study (eg, one email that had 1 reinforcing comment and 1 area for change, one email that had 4 reinforcing comments and 8 areas for change). These emails were reviewed by 3 experienced interventionists (email text is presented in Multimedia Appendix 1).

Measures

Respondents first reported their demographics (ie, age, gender, race, and ethnicity) and socioeconomic characteristics (ie, annual household income, highest level of education, and current employment status) using questions from the Behavioral Risk Factor Surveillance System [18].

Likelihood to self-monitor in the future was measured using 1 item ("How likely is it that you would continue to record your eating and physical activity in the online diary in the next week?") with a 5-point Likert response scale (1=*very unlikely* to 5=*very likely*).

Motivation to make behavioral changes in the future was measured using 1 item ("How motivated would you be to continue to make behavioral changes in your eating and physical activity in the next week?") with a 5-point Likert response scale (1=*not motivated* to 5=*very motivated*).

The perception of the counselor included 3 components, each measured by 1 item. The first 2 items were adapted from the Working Alliance Inventory Bond subscale [19]: "The counselor is genuinely concerned for my welfare" and "The counselor genuinely likes me." Participants also responded to the following statement: "The counselor is understanding about my challenges." Participants' responses were recorded on a 5-point Likert scale (1=*strongly disagree* to 5=*strongly agree*).

The survey also captured impressions about 3 aspects of the emails, with items that inquired about the insightfulness of the feedback ("The e-mail gives me a new way of looking at weight and eating and physical activity behaviors," adapted from the Working Alliance Inventory [19]), how tailored they perceived the email to be ("The information in the e-mail seems tailored," adapted from a study by Valle et al [20]), and perceptions about the length of email ("I feel that the e-mail is... 'too short' to 'too long'"). All items used a 5-point Likert response format.

Independent Variables

The independent variables indicated the level of reinforcing comments (ie, 1 comment, 4 comments, and 8 comments) and the level of areas for change (ie, 1 area, 4 areas, and 8 areas). Interaction terms between the levels of reinforcing comments and areas for change were constructed to assess whether respondents' likelihood to self-monitor and make behavioral changes in the future and their perceptions of the counselor and the email would change, given different combinations of feedback.

Analyses

Demographic and socioeconomic characteristics were tabulated to examine the sample attributes. As each individual responded to all 9 email variations, mixed effects ordinal regression was used to model the 5-level ordinal responses. A random intercept for each individual was used. The model included the main effects of reinforcing comments, areas of change, and all two-factor interactions on respondents' likelihood to self-monitor, motivation to make behavioral changes, and perceptions of the counselor and emails. The reference category for reinforcing comments is 1 reinforcing comment, and the reference category for areas for change is 1 area for change.

To effectively assess individuals' perception of the self-monitoring feedback emails as well as for ease of interpretability, we calculated average predicted probabilities of the most positive category (ie, very likely, strongly agree, very motivated, and just right), conditional on the value of the random effect when the interaction between the 2 independent variables was included. Given that sociodemographic characteristics have been shown to affect self-monitoring rates [21-24], we adjusted the model to take into account age, ethnicity, race, gender, income, education level, employment status, and BMI. Analyses were conducted using STATA (version 16).

Results

Sample Characteristics

The demographic and socioeconomic characteristics of the participants are reported in Table 1. Approximately equal numbers of women and men participated, and most participants had at least some college education (74.0%). The sample was approximately evenly distributed between individuals with overweight and obesity.

Table 1. Sample sociodemographic characteristics (N=250).

Sociodemographic characteristics	Values, n (%)
Age (years)	
18-24	10 (4.0)
25-34	31 (12.4)
35-44	46 (18.4)
45-54	31 (12.4)
55-64	33 (13.2)
≥65	97 (38.8)
Missing	2 (0.8)
Gender	
Women	130 (52.0)
Men	120 (48.0)
Race	
White	127 (50.8)
Black or African American	93 (37.2)
Other	30 (12.0)
Ethnicity	
Hispanic or Latino	23 (9.2)
Educational level	
Lower than high school degree	64 (25.6)
Some college	80 (32.0)
Higher than college degree	105 (42.0)
Missing	1 (0.4)
Employed	115 (46.0)
Annual household income (US $)	
≤24,999	54 (21.6)
25,000-49, 999	73 (29.2)
50,000-74,999	47 (18.8)
≥75,000	75 (30.0)
Missing	1 (0.4)
BMI category (kg/m^2)	
Obese (BMI≥30)	137 (54.8)
Overweight (BMI=25.0-29.9)	113 (45.2)

Likelihood to Self-Monitor in the Future

The odds that an individual expressed the likelihood to self-monitor in the future was at least 3 times higher after reading emails with 4 (odds ratio [OR] 3.22, 95% CI 2.18-4.75) or 8 reinforcing comments (OR 3.75, 95% CI 2.54-5.54) than after reading emails with 1 reinforcing comment (Table 2). In addition, the odds that individuals expressed a likelihood to self-monitor in the future was at least 2 times higher after reading emails with 4 (OR 2.24, 95% CI 1.52-3.28) or 8 areas for change (OR 2.58, 95% CI 1.75-3.80) than after reading emails with 1 area for change. The predictive margins on the interaction terms demonstrated that individuals reported the highest probability of being very likely to self-monitor in the future after reading the email with 4 reinforcing comments and 4 areas for change (49%, 95% CI 0.43-0.54; Table 3). Similar findings were observed with regard to the likelihood to self-monitor in the future based on levels of reinforcing comment, areas for change, and the interactions in the models that adjusted for sociodemographic characteristics and BMI category (Table 4).

Table 2. Odds ratios (with 95% CI) of different combinations of reinforcing comments and areas for change on all outcome measures (N=250).

Feedback content	Motivation, OR[a,b] (95% CI)		Perception of the counselor, OR (95% CI)			Perception of the email, OR (95% CI)		
	Self-monitor in the future	Motivated to make behavior changes	Counselor concerned about welfare	Counselor likes me	Counselor understands challenges	Email gives insight	Email is tailored	Email length
Level of reinforcing comments (reference=1 reinforcing comment)								
4 reinforcing comments	3.22 (2.18-4.75)	3.27 (2.24-4.78)	4.26 (2.91-6.22)	2.55 (1.73-3.76)	3.80 (2.62-5.50)	4.69 (3.23-6.83)	2.72 (1.89-3.92)	10.46 (7.01-15.62)
8 reinforcing comments	3.75 (2.54-5.54)	4.76 (3.24-7.00)	4.98 (3.39-7.31)	3.11 (2.11-4.60)	5.32 (3.64-7.77)	5.70 (3.90-8.35)	3.50 (2.41-5.08)	17.65 (11.70-26.63)
Level of areas for change (reference=1 area for change)								
4 areas for change	2.24 (1.52-3.28)	2.27 (1.55-3.31)	3.13 (2.15-4.56)	2.34 (1.59-3.46)	3.12 (2.16-4.52)	3.81 (2.63-5.54)	2.62 (1.81-3.79)	13.61 (9.08-20.39)
8 areas for change	2.58 (1.75-3.80)	2.52 (1.73-3.68)	4.61 (3.15-6.74)	2.50 (1.70-3.69)	4.45 (3.06-6.47)	5.01 (3.44-7.29)	3.40 (2.35-4.92)	107.55 (69.41-166.63)
Interactions								
4 reinforcing comments and 4 areas for change	0.58 (0.34-1.00)	0.57 (0.33-0.96)	0.37 (0.22-0.63)	0.54 (0.32-0.93)	0.49 (0.29-0.83)	0.33 (0.20-0.56)	0.39 (0.23-0.64)	0.13 (0.08-0.23)
4 reinforcing comments and 8 areas for change	0.38 (0.22-0.66)	0.45 (0.27-0.77)	0.29 (0.17-0.49)	0.55 (0.32-0.95)	0.30 (0.18-0.50)	0.27 (0.16-0.46)	0.34 (0.20-0.57)	0.11 (0.07-0.19)
8 reinforcing comments and 4 areas for change	0.44 (0.26-0.76)	0.44 (0.26-0.75)	0.48 (0.28-0.81)	0.54 (0.31-0.92)	0.30 (0.18-0.51)	0.29 (0.17-0.48)	0.36 (0.22-0.61)	0.24 (0.14-0.40)
8 reinforcing comments and 8 areas for change	0.25 (0.15-0.43)	0.24 (0.14-0.41)	0.28 (0.16-0.47)	0.33 (0.19-0.58)	0.16 (0.10-0.28)	0.21 (0.12-0.36)	0.30 (0.18-0.51)	0.14 (0.08-0.23)

[a]OR: odds ratio.

[b]All odds ratios are significant at *P*<.05.

Table 3. Average predicted probabilities of combinations of email feedback on all outcome measures (N=250).[a]

Feedback content	Motivation, APP[b] (95% CI)		Perception of the counselor, APP (95% CI)			Perception of the email, APP (95% CI)		
	Very likely to self-monitor in the future	Very likely to be motivated to make behavior changes	Strongly agree that the counselor is concerned about welfare	Strongly agree the counselor likes me	Strongly agree the counselor understands challenges	Strongly agree the email gives insight	Strongly agree the email is tailored	Length of the email is just right
1 reinforcing comment and 1 area for change	0.32 (0.27-0.36)	0.27 (0.23-0.32)	0.21 (0.17-0.25)	0.20 (0.16-0.24)	0.21 (0.17-0.26)	0.19 (0.15-0.24)	0.25 (0.21-0.30)	0.43 (0.37-0.48)
1 reinforcing comment and 4 areas for change	0.41 (0.35-0.46)	0.36 (0.31-0.41)	0.32 (0.27-0.37)	0.27 (0.23-0.32)	0.34 (0.29-0.39)	0.33 (0.29-0.38)	0.35 (0.30-0.40)	*0.65*[c] (0.62-0.68)
1 reinforcing comment and 8 areas for change	0.42 (0.37-0.48)	0.37 (0.32-0.42)	0.36 (0.31-0.41)	0.28 (0.23-0.32)	0.38 (0.33-0.43)	0.37 (0.32-0.42)	*0.39*[c] (0.33-0.44)	0.45 (0.40-0.50)
4 reinforcing comments and 1 area for change	0.45 (0.39-0.51)	0.40 (0.35-0.45)	0.35 (0.31-0.40)	0.28 (0.24-0.32)	0.36 (0.31-0.41)	0.36 (0.31-0.41)	0.36 (0.31-0.41)	*0.65*[c] (0.62-0.68)
4 reinforcing comments and 4 areas for change	*0.49*[c] (0.43-0.54)	0.43 (0.37-0.49)	0.37 (0.32-0.42)	0.30 (0.26-0.35)	*0.42*[c] (0.36-0.47)	0.39 (0.34-0.44)	0.36 (0.31-0.41)	0.64 (0.60-0.67)
4 reinforcing comments and 8 areas for change	0.45 (0.39-0.51)	0.42 (0.36-0.47)	0.39 (0.34-0.44)	0.31 (0.27-0.36)	0.34 (0.35-0.45)	*0.40*[c] (0.35-0.45)	0.38 (0.32-0.43)	0.43 (0.38-0.48)
8 reinforcing comments and 1 area for change	0.47 (0.41-0.53)	*0.45*[c] (0.39-0.50)	0.37 (0.32-0.43)	0.30 (0.25-0.34)	0.41 (0.35-0.46)	0.38 (0.33-0.44)	*0.39*[c] (0.34-0.44)	0.64 (0.61-0.67)
8 reinforcing comments and 4 areas for change	0.47 (0.41-0.53)	*0.45*[c] (0.39-0.50)	*0.42*[c] (0.37-0.48)	*0.32*[c] (0.27-0.37)	0.40 (0.34-0.45)	0.39 (0.34-0.45)	0.38 (0.33-0.44)	0.54 (0.50-0.58)
8 reinforcing comments and 8 areas for change	0.42 (0.36-0.47)	0.38 (0.33-0.44)	0.40 (0.35-0.46)	0.28 (0.24-0.33)	0.36 (0.31-0.41)	0.39 (0.34-0.44)	*0.39*[c] (0.34-0.45)	0.32 (0.28-0.37)

[a]All probabilities were significant at the $P<.01$ level.

[b]APP: average predicted probabilities.

[c]Highest probability (or probabilities) is italicized.

Table 4. Adjusted odds ratios for sociodemographic factors on all outcome measures (N=250).[a]

Characteristics	Motivation, OR[b] (95% CI)		Perception of the counselor, OR (95% CI)			Perception of the email, OR (95% CI)		
	Self-monitor in the future	Motivated to make behavior changes	Counselor concerned about welfare	Counselor likes me	Counselor understands challenges	Email gives insight	Email is tailored	Email length
Level of reinforcing comments (reference=1)								
4 reinforcing comments	3.22[***][c] (2.18-4.77)	3.34[***] (2.28-4.91)	4.33[***] (2.96-6.35)	2.50[***] (1.69-3.70)	3.76[***] (2.59-5.47)	4.68[***] (3.21-6.82)	2.78[***] (1.92-4.02)	10.64[***] (7.10-15.94)
8 reinforcing comments	3.87[***] (2.61-5.73)	4.74[***] (3.21-6.99)	5.03[***] (3.42-7.41)	3.22[***] (2.17-4.78)	5.57[***] (3.79-8.18)	5.75[***] (3.92-8.45)	3.68[***] (2.52-5.36)	18.26[***] (12.04-27.68)
Level of areas for change (reference=1)								
4 areas for change	2.25[***] (1.53-3.32)	2.28[***] (1.56-3.34)	3.08[***] (2.12-4.50)	2.27[***] (1.54-3.36)	3.16[***] (2.18-4.58)	3.82[***] (2.62-5.56)	2.62[***] (1.80-3.80)	13.95[***] (9.27-21.01)
8 areas for change	2.52[***] (1.71-3.72)	2.53[***] (1.73-3.70)	4.58[***] (3.13-6.72)	2.54[***] (1.72-3.76)	4.59[***] (3.14-6.69)	5.04[***] (3.45-7.36)	3.52[***] (2.42-5.10)	114.68[***] (73.61-178.65)
Interactions								
4 reinforcing comments and 4 areas for change	0.59[*] (0.34-1.01)	0.56[**] (0.33-0.95)	0.38[***] (0.22-0.64)	0.58[*] (0.34-1.01)	0.51[**] (0.30-0.85)	0.34[***] (0.20-0.56)	0.39[***] (0.24-0.66)	0.14[***] (0.08-0.24)
4 reinforcing comments and 8 areas for change	0.40[***] (0.23-0.69)	0.45[***] (0.26-0.77)	0.29[***] (0.17-0.49)	0.57[**] (0.33-0.98)	0.30[***] (0.18-0.50)	0.28[***] (0.17-0.47)	0.33[***] (0.20-0.56)	0.11[***] (0.06-0.19)
8 reinforcing comments and 4 areas for change	0.43[***] (0.25-0.75)	0.44[***] (0.26-0.76)	0.48[***] (0.28-0.82)	0.56[**] (0.32-0.96)	0.30[***] (0.18-0.51)	0.29[***] (0.17-0.49)	0.35[***] (0.21-0.59)	0.24[***] (0.14-0.42)
8 reinforcing comments and 8 areas for change	0.24[***] (0.14-0.42)	0.23[***] (0.14-0.40)	0.27[***] (0.16-0.46)	0.31[***] (0.18-0.54)	0.15[***] (0.09-0.26)	0.21[***] (0.12-0.36)	0.28[***] (0.16-0.47)	0.13[***] (0.08-0.23)
Sociodemographic characteristics								
Age (reference: 18-24 years), years								
25-34	11.96[**] (1.46-98.04)	13.21[**] (1.82-95.95)	18.40[***] (2.71-124.84)	16.51[***] (1.99-136.78)	7.25[**] (1.26-41.73)	10.30[**] (1.72-61.56)	9.56[**] (1.43-63.72)	0.77 (0.22-2.75)
35-44	7.65[*] (1.00-58.65)	11.83[**] (1.73-80.80)	17.33[***] (2.71-110.77)	10.70[**] (1.38-82.97)	6.46[**] (1.19-35.18)	6.25[**] (1.11-35.27)	3.17 (0.51-19.82)	1.36 (0.40-4.67)
45-54	3.16 (0.39-25.66)	6.80[*] (0.94-49.01)	17.72[***] (2.63-119.48)	6.53[*] (0.80-53.58)	5.25[*] (0.92-30.07)	5.62[*] (0.95-33.32)	5.03[*] (0.76-33.22)	0.74 (0.21-2.63)
55-64	4.33 (0.53-35.71)	8.42[**] (1.15-61.68)	29.52[***] (4.30-202.73)	7.66[*] (0.92-63.79)	9.05[**] (1.56-52.46)	13.14[***] (2.17-79.42)	5.79[*] (0.87-38.75)	1.04 (0.29-3.73)
≥65	4.62 (0.62-34.41)	9.10[**] (1.37-60.50)	18.46[***] (2.96-115.02)	6.82[*] (0.90-51.46)	6.46[**] (1.22-34.33)	9.17[**] (1.66-50.53)	2.66 (0.44-16.18)	0.87 (0.26-2.94)
Hispanic or Latino (reference: Non-Hispanic)	1.54 (0.42-5.72)	1.45 (0.42-5.01)	1.349 (0.41-4.45)	1.40 (0.38-5.16)	0.88 (0.30-2.62)	1.15 (0.38-3.53)	1.01 (0.31-3.30)	0.50[*] (0.23-1.11)
Race (reference: White)								

XSL•FO
RenderX

Characteristics	Motivation, OR[b] (95% CI)		Perception of the counselor, OR (95% CI)			Perception of the email, OR (95% CI)		
	Self-monitor in the future	Motivated to make behavior changes	Counselor concerned about welfare	Counselor likes me	Counselor understands challenges	Email gives insight	Email is tailored	Email length
Black or African American	4.34[***] (1.79-10.51)	4.05[***] (1.75-9.39)	2.32[**] (1.03-5.23)	1.79 (0.74-4.32)	2.06[*] (0.98-4.32)	3.00[***] (1.40-6.43)	1.02 (0.46-2.27)	0.89 (0.52-1.53)
Other	0.74 (0.21-2.62)	1.166 (0.35-3.85)	0.99 (0.31-3.13)	0.79 (0.22-2.77)	1.07 (0.38-3.07)	1.40 (0.47-4.11)	0.89 (0.28-2.79)	1.13 (0.52-2.42)
Gender (reference: men)								
Women	1.49 (0.70-3.16)	2.00[*] (0.98-4.08)	2.32[**] (1.17-4.62)	2.09[*] (0.98-4.42)	2.38[***] (1.27-4.48)	2.39[***] (1.25-4.55)	1.86[*] (0.94-3.69)	0.89 (0.56-1.40)
Annual household income (reference: ≤24,999), US $								
25,000-49,999	1.19 (0.43-3.27)	1.15 (0.44-3.01)	0.86 (0.34-2.17)	1.05 (0.38-2.92)	0.86 (0.37-2.00)	0.89 (0.37-2.12)	0.56 (0.22-1.41)	1.35 (0.73-2.49)
50,000-74,999	2.74[*] (0.84-9.01)	1.61 (0.52-4.97)	1.34 (0.45-3.95)	1.49 (0.45-4.89)	1.09 (0.41-2.93)	1.70 (0.62-4.68)	1.03 (0.35-3.03)	1.318 (0.64-2.71)
≥75,000	3.73[**] (1.22-11.43)	2.32 (0.80-6.72)	2.70[*] (0.97-7.52)	2.34 (0.76-7.17)	2.53[*] (0.99-6.44)	2.52[*] (0.97-6.56)	1.93 (0.70-5.34)	0.93 (0.47-1.83)
Educational level (reference: high school degree or less)								
Some college	0.46 (0.18-1.17)	0.64 (0.26-1.54)	0.76 (0.33-1.79)	0.56 (0.22-1.41)	0.55 (0.25-1.19)	0.59 (0.27-1.32)	1.10 (0.47-2.56)	0.79 (0.45-1.39)
College degree or more	0.78 (0.30-2.04)	0.89 (0.36-2.23)	1.04 (0.43-2.51)	0.54 (0.21-1.42)	0.74 (0.33-1.65)	0.67 (0.29-1.54)	1.15 (0.48-2.76)	0.58[*] (0.32-1.04)
Employment status (reference: not employed)								
Employed	0.61 (0.24-1.53)	0.67 (0.28-1.61)	1.26 (0.54-2.95)	2.22[*] (0.88-5.61)	1.46 (0.67-3.16)	1.40 (0.64-3.10)	1.27 (0.55-2.93)	1.25 (0.71-2.18)
BMI category (reference: overweight)								
Obese	0.50[*] (0.24-1.04)	0.42[**] (0.21-0.85)	0.60 (0.31-1.18)	0.81 (0.39-1.69)	0.62 (0.33-1.14)	0.72 (0.38-1.34)	0.63 (0.33-1.23)	0.83 (0.53-1.27)

[a]Adjusted for sociodemographic characteristics (ie, age, ethnicity, race, gender, income, education level, employment status, and BMI category).

[b]OR: odds ratio.

[c]*$P<.1$, **$P<.05$, ***$P<.01$.

Overall, individuals with certain sociodemographic characteristics had significantly higher odds of reporting a likelihood to self-monitor in the future (Table 4); these characteristics were identifying as Black or African American (OR 4.34, 95% CI 1.79-10.51), having an annual household income of US $50,000 to US $74,999 (OR 2.74, 95% CI 0.84-9.01) or US $75,000 or more (OR 3.73, 95% CI 1.22-11.43), and aged 25 to 44 years (OR 7.65-11.96). In addition, those with obesity (OR 0.50, 95% CI 0.24-1.04) were less likely to report a likelihood to self-monitor in the future, compared with those with overweight.

Motivation to Make Behavioral Changes

The odds that individuals would be motivated to make behavioral changes was 3 to almost 5 times higher after reading emails with 4 (OR 3.27, 95% CI 2.24-4.78) or 8 (OR 4.76, 95% CI 3.24-7.00) reinforcing comments than after reading emails with 1 reinforcing comment, and the odds that individuals would be motivated to make behavioral changes was at least 2 times

higher after reading emails with 4 (OR 2.27, 95% CI 1.55-3.31) or 8 (OR 2.52, 95% CI 1.73-3.68) areas for change (Table 2). The predictive margins on the interaction terms demonstrated that individuals reported the highest probability of being very likely to be motivated to make behavioral changes after reading the email with 8 reinforcing comments and 1 area for change (45%; 95% CI 0.39-0.50; Table 3) or the email with 8 reinforcing comments and 4 areas for change (45%; 95% CI 0.39-0.50).

Again, there were similar outcomes in the models that adjusted for sociodemographic characteristics and BMI category (Table 4). Those who identified as Black or African American (OR 4.05, 95% CI 1.75-9.39) or as a woman (OR 2.00, 95% CI 0.98-4.08) were more likely to report being motivated to make behavioral changes (Table 4). Respondents older than 24 years had significantly higher odds to report being motivated to make behavioral changes (ORs 6.80-13.21) than those aged 18-24 years. Respondents with obesity (OR 0.42, 95% CI 0.21-0.85)

were less likely to report being motivated to make behavioral changes compared with those who were overweight.

Perception of Counselor

The odds that individuals rated the counselor as concerned for their welfare was 4-5 times higher after reading emails with 4 (OR 4.26, 95% CI 2.91-6.22) or 8 (OR 4.98, 95% CI 3.39-7.31) reinforcing comments than after reading those with 1 reinforcing comment, and 3-4 times higher after reading emails with 4 (OR 3.13, 95% CI 2.15-4.56) or 8 (OR 4.61, 95% CI 3.15-6.74) areas for change (Table 2) than after reading emails with 1 area for change. In addition, the odds that individuals indicated that the counselor genuinely liked them were significantly higher after reading emails with 4 (OR 2.55, 95% CI 1.73-3.76) or 8 (OR 3.11, 95% CI 2.11-4.60) reinforcing comments or 4 (OR 2.34, 95% CI 1.59-3.46) or 8 (OR 2.50, 95% CI 1.70-3.69) areas for change than after reading emails with only one reinforcing comment or 1 area for change. Furthermore, the odds that individuals reported that the counselor understands their challenges was significantly higher after reading emails with 4 (OR 3.80, 95% CI 2.62-5.50) or 8 (OR 5.32, 95% CI 3.64-7.77) reinforcing comments or 4 (OR 3.12, 95% CI 2.16-4.52) or 8 (OR 4.45, 95% CI 3.06-6.47) areas for change.

The interactions indicated that individuals had the highest average probability of strongly agreeing that the counselor was genuinely concerned for their welfare (42%; 95% CI 0.37-0.48) after reading emails with 8 reinforcing comments and 4 areas for change (Table 3). In addition, individuals had the highest average probability of strongly agreeing with the counselor genuinely liked them (32%; 95% CI 0.27-0.37) after reading emails with 8 reinforcing comments and 4 areas for change. As for the perception of the counselor being understanding of their challenges, the interactions demonstrated that individuals had the highest average probability of strongly agreeing after reading emails with 4 reinforcing comments and 4 areas for change (42%; 95% CI 0.36-0.47). The sociodemographic group differences in perceptions of the counselor were similar to those described earlier (Table 4).

Perception of the Emails

Perceptions of insightfulness, tailoring, and length of the emails were significantly more favorable after reading emails with 4 or 8 reinforcing comments or 4 or 8 areas for change than after reading emails with only one reinforcing comment or 1 area for change (Table 2). The interactions demonstrated that individuals had the highest average probability of strongly agreeing that the emails gave them a new way of looking at diet and physical activity behaviors (40%; 95% CI 0.35-0.45) after reading emails with 4 reinforcing comments and 8 areas for change (Table 3). Perceptions that the emails were tailored were strongest (39%; 95% CI 0.33-0.44) after reading emails with 1 reinforcing comment and 8 areas for change, 8 reinforcing comments and 1 area for change, and 8 reinforcing comments and 8 areas for change. Finally, perceptions that the length of the emails being just right were the highest (65%; 95% CI 0.62-0.68) after reading emails with 4 reinforcing comments and 1 area for change and emails with 1 reinforcing comment and 4 areas for change. The sociodemographic differences in the perception of emails are similar to those described earlier (Table 4).

Discussion

This study conducted a factorial experiment to examine the perception of email feedback on diet and physical activity behaviors among individuals with overweight or obesity and a desire to lose weight. The main effects showed that the emails with more reinforcing comments or areas for change were better received, in general, with often a small incremental benefit for 8 reinforcing comments or areas for change versus 4 reinforcing comments or areas for change. Thus, participants may view longer messages that provide the overall context of their behaviors over a week as more positive, rather than shorter messages containing just one reinforcing comment or area for change. The interactions indicated that the best combination for 3 of 8 dimensions assessed was the email with 8 reinforcing comments and 4 areas for change, which was associated with higher ratings of motivation to make behavioral changes in the future, counselor's concern for their welfare, and perceptions that the counselor genuinely likes them. Nonetheless, as consistent self-monitoring is important for effective weight management [12], perhaps the best combination among those tested in this study is the email with 4 reinforcing comments and 4 areas for change because it resulted in the highest average probability of reporting being very likely to self-monitor in the future if continuing to self-monitor is weighted most heavily.

Interestingly, the longest email combinations (ie, the 8 reinforcing comments and 8 areas for change, the 1 reinforcing comment and 8 areas for change, and the 8 reinforcing comments and 1 area for change combinations) had the highest average probability of being rated as being tailored. Perhaps, individuals viewed more feedback as being more customized to their situation. In contrast, the email combination that had the maximum number of areas of change (ie, 8) and a moderate number of reinforcing comments (ie, 4) had the highest likelihood of being viewed as insightful, indicating that insightfulness might be perceived as a balance between reinforcing comments and areas for change. However, the emails judged as *just right* in length were shorter than any of these options (ie, the 1 reinforcing comment and 4 areas for change and the 4 reinforcing comments and 1 area for change combinations). Thus, the optimal constellation of areas of change and number of reinforcing comments may depend on the outcome, and there are few data to guide which of these outcomes is associated with consistent self-monitoring and, ultimately, superior weight loss.

The responses to the different email feedback combinations varied very little across the demographic groups considered, which is interesting as some groups have been found to have significantly higher or lower self-monitoring engagement in previous studies. For example, studies have found that women have lower engagement with self-monitoring than men [21,22], whereas we found that women and men in this study reported a similar likelihood to self-monitor in the future, although it must be acknowledged that these were hypothetical projections rather than actual self-monitoring behavior. In addition, previous research indicates that older individuals are more likely to engage in self-monitoring than younger individuals [21,23,24], although we found that individuals aged 25 to 44 years had a

higher odds to report that they were likely to self-monitor in the future than individuals who were aged 18 to 24 years and the older age groups did not differ from the youngest group. This is in contrast to previous research, which found that older adults (ie, aged 60-85 years) were more successful in losing weight in the Diabetes Prevention Program's weight management intervention [25], perhaps because older individuals have fewer competing obligations. Finally, although Black respondents indicated that they would be more likely to self-monitor in the future, previous research has not found significant race-based differences in self-monitoring engagement [17,21]. Nonetheless, a meta-analysis that examined the impact of tailored messages for health behavioral change in general also found a minimal impact of sociodemographic factors on the effect of the messages [5]. It will be important to examine whether responses to the different email feedback combinations are comparable across demographic groups when studying actual self-monitoring behavior, rather than self-reported likelihood of self-monitoring in the future.

There are some limitations to this study, which merit acknowledgment. Importantly, participants who rated the emails were not enrolled in a weight loss study and thus provided hypothetical responses to feedback emails that were not personalized to the individual in this study. It is unknown if the likelihood of self-monitoring or motivation to make behavioral changes in the future would translate to actual behavior. However, all respondents were either overweight or obese and expressed an interest in losing weight, so it may be assumed that the sample resembled respondents who would join a weight loss program to some extent. In addition, the concepts were each measured by a single item. Although these single items were adapted from previous studies when possible and have face validity, the psychometric properties of the items were not examined. Future studies should examine individuals actively engaged in weight loss efforts to determine whether the composition of emailed feedback drives actual self-monitoring

patterns and ultimately influences weight loss outcomes. Furthermore, we arbitrarily chose the 3 levels of each variable (ie, 1, 4, and 8) so that there was separation between each level and to represent what might be considered a small number of comments, a moderate number of comments, and a large number of comments. Furthermore, we tested only 3 levels of each variable, and thus, it is possible that the *ideal* combination was not examined in this study. In future research, it will be important to test other levels of the variables, including 0 reinforcing comments or areas for change; however, researchers need to be aware of participant burden and attention limitations, which is why we limited our survey to 9 emails. Despite these limitations, to our knowledge, this is the first study to explore perspectives on self-monitoring feedback and to provide the first indication of the *sweet spots* in providing weekly self-monitoring feedback. An additional strength is the diversity of the sample with respect to gender and race and ethnicity, which allowed some insights along those dimensions.

Given the broad and escalating need for obesity treatment, it is essential to optimize treatment components for efficacy and cost-effectiveness, including self-monitoring feedback. A recent study reported that the average time for counselors to review self-monitoring records and compose weekly dietary and physical activity self-monitoring feedback emails is approximately 28 min, following a guideline of 6 reinforcing comments and 3 areas for change [26]. Some have opted to use computer-driven algorithms to generate feedback [27], whereas others report that human-generated feedback produces superior weight loss outcomes [4]. It will be important in future research to examine whether the optimal approach for self-monitoring feedback differs in the context of human-generated emails compared with algorithm-driven emails and to identify the most time-efficient (and therefore least costly) feedback constellation for humans to generate. This study offers some specific hypotheses to test and, as such, can advance this critical aspect of behavioral weight management intervention refinement.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 9 emails used in this research. Blue text indicates the combination of the number of reinforcing comments and areas for change. Red text indicates an area for change, and green text indicates a reinforcing comment. Black text is standard for every email.
[DOCX File , 34 KB - jmir_v22i11e18104_app1.docx]

References

1. Steinberg DM, Levine EL, Askew S, Foley P, Bennett GG. Daily text messaging for weight control among racial and ethnic minority women: randomized controlled pilot study. J Med Internet Res 2013 Nov 18;15(11):e244 [FREE Full text] [doi: 10.2196/jmir.2844] [Medline: 24246427]
2. Diabetes Prevention Program (DPP) Research Group. The diabetes prevention program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]

3. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]

4. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]

5. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]

6. West DS, Bursac Z, Cornell CE, Felix HC, Fausett JK, Krukowski RA, et al. Lay health educators translate a weight-loss intervention in senior centers: a randomized controlled trial. Am J Prev Med 2011 Oct;41(4):385-391 [FREE Full text] [doi: 10.1016/j.amepre.2011.06.041] [Medline: 21961465]

7. Resnicow K, McMaster F. Motivational interviewing: moving from why to how with autonomy support. Int J Behav Nutr Phys Act 2012 Mar 2;9:19 [FREE Full text] [doi: 10.1186/1479-5868-9-19] [Medline: 22385702]

8. Ritzwoller DP, Glasgow RE, Sukhanova AY, Bennett GG, Warner ET, Greaney ML, Be Fit Be Well Study Investigators. Economic analyses of the Be Fit Be Well program: a weight loss program for community health centers. J Gen Intern Med 2013 Dec;28(12):1581-1588. [doi: 10.1007/s11606-013-2492-3] [Medline: 23733374]

9. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]

10. Michie S. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008 Feb 1;57:660-680. [doi: 10.1111/j.1464-0597]

11. Schunk DH, Swartz CW. Goals and progress feedback: effects on self-efficacy and writing achievement. Contemp Educ Psychol 1993 Jul;18(3):337-354. [doi: 10.1006/ceps.1993.1024]

12. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]

13. Silva MN, Markland D, Minderico CS, Vieira PN, Castro MM, Coutinho SR, et al. A randomized controlled trial to evaluate self-determination theory for exercise adherence and weight control: rationale and intervention description. BMC Public Health 2008 Jul 9;8:234 [FREE Full text] [doi: 10.1186/1471-2458-8-234] [Medline: 18613959]

14. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. J Pers Soc Psychol 1996 Jan;70(1):115-126. [doi: 10.1037//0022-3514.70.1.115] [Medline: 8558405]

15. Gorin AA, Powers TA, Koestner R, Wing RR, Raynor HA. Autonomy support, self-regulation, and weight loss. Health Psychol 2014 Apr;33(4):332-339 [FREE Full text] [doi: 10.1037/a0032586] [Medline: 23730718]

16. Unlock Breakthrough Insights With Market Research Panels. Qualtrics XM: The Leading Experience Management Software. 2019. URL: https://www.qualtrics.com/online-sample/ [accessed 2019-04-01]

17. West DS, Krukowski RA, Finkelstein EA, Stansbury ML, Ogden DE, Monroe CM, et al. Adding financial incentives to online group-based behavioral weight control: an RCT. Am J Prev Med 2020 Aug;59(2):237-246. [doi: 10.1016/j.amepre.2020.03.015] [Medline: 32446752]

18. BRFSS 2012 Survey Data and Documentation. Centers for Disease Control and Prevention. 2014 Nov 12. URL: http://www.cdc.gov/brfss/annual_data/annual_2012.html [accessed 2020-09-14]

19. Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. J Counsel Psychol 1989;36(2):223. [doi: 10.1037/0022-0167.36]

20. Valle CG, Queen TL, Martin BA, Ribisl KM, Mayer DK, Tate DF. Optimizing tailored communications for health risk assessment: a randomized factorial experiment of the effects of expectancy priming, autonomy support, and exemplification. J Med Internet Res 2018 Mar 1;20(3):e63 [FREE Full text] [doi: 10.2196/jmir.7613] [Medline: 29496652]

21. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170 [FREE Full text] [doi: 10.1037/a0028135] [Medline: 22545978]

22. Hollis JF, Gullion CM, Stevens VJ, Brantley PJ, Appel LJ, Ard JD, Weight Loss Maintenance Trial Research Group. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. Am J Prev Med 2008 Aug;35(2):118-126 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.013] [Medline: 18617080]

23. Mossavar-Rahmani Y, Henry H, Rodabough R, Bragg C, Brewer A, Freed T, et al. Additional self-monitoring tools in the dietary modification component of the women's health initiative. J Am Diet Assoc 2004 Jan;104(1):76-85. [doi: 10.1016/j.jada.2003.10.017] [Medline: 14702588]

24. Sevick MA, Stone RA, Zickmund S, Wang Y, Korytkowski M, Burke LE. Factors associated with probability of personal digital assistant-based dietary self-monitoring in those with type 2 diabetes. J Behav Med 2010 Aug;33(4):315-325. [doi: 10.1007/s10865-010-9257-9] [Medline: 20232131]

25. Diabetes Prevention Program Research Group, Crandall J, Schade D, Ma Y, Fujimoto WY, Barrett-Connor E, et al. The influence of age on the effects of lifestyle modification and metformin in prevention of diabetes. J Gerontol A Biol Sci Med Sci 2006 Oct;61(10):1075-1081 [FREE Full text] [doi: 10.1093/gerona/61.10.1075] [Medline: 17077202]

26.   Fahey M, Kocak M, Gladney L, Talcott G, Krukowski R. Do counselors become more efficient at providing self-monitoring feedback in a technology-based behavioral weight loss intervention? JMIR Preprints Preprint posted online September 5, 2020. [doi: 10.2196/preprints.23974]
27.   Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mhealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]

Abbreviations

OR: odds ratio

XSL•FO
RenderX

Original Paper

Acceptability of Using a Robotic Nursing Assistant in Health Care Environments: Experimental Pilot Study

Mohammad Nasser Saadatzi[1], PhD; M Cynthia Logsdon[2], PhD; Shamsudeen Abubakar[1], MSc; Sumit Das[1], PhD; Penelope Jankoski[3], BSc; Heather Mitchell[2], PhD; Diane Chlebowy[2], PhD; Dan O Popa[1], PhD

[1]Electrical and Computer Engineering Department, University of Louisville, Louisville, KY, United States
[2]School of Nursing, University of Louisville, Louisville, KY, United States
[3]Biomedical Engineering Department, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Mohammad Nasser Saadatzi, PhD
Electrical and Computer Engineering Department
University of Louisville
2210 S Brook St
Louisville, KY, 40208
United States
Phone: 1 502 852 5555
Email: nasser_saadatzi@yahoo.com

Abstract

Background: According to the US Bureau of Labor Statistics, nurses will be the largest labor pool in the United States by 2022, and more than 1.1 million nursing positions have to be filled by then in order to avoid a nursing shortage. In addition, the incidence rate of musculoskeletal disorders in nurses is above average in comparison with other occupations. Robot-assisted health care has the potential to alleviate the nursing shortage by automating mundane and routine nursing tasks. Furthermore, robots in health care environments may assist with safe patient mobility and handling and may thereby reduce the likelihood of musculoskeletal disorders.

Objective: This pilot study investigates the perceived ease of use and perceived usefulness (acceptability) of a customized service robot as determined by nursing students (as proxies for nursing staff in health care environments). This service robot, referred to as the Adaptive Robotic Nurse Assistant (ARNA), was developed to enhance the productivity of nurses through cooperation during physical tasks (eg, patient walking, item fetching, object delivery) as well as nonphysical tasks (eg, patient observation and feedback). This pilot study evaluated the acceptability of ARNA to provide ambulatory assistance to patients.

Methods: We conducted a trial with 24 participants to collect data and address the following research question: Is the use of ARNA as an ambulatory assistive device for patients acceptable to nurses? The experiments were conducted in a simulated hospital environment. Nursing students (as proxies for nursing staff) were grouped in dyads, with one participant serving as a nurse and the other acting as a patient. Two questionnaires were developed and administrated to the participants based on the Technology Acceptance Model with respect to the two subscales of perceived usefulness and perceived ease of use metrics. In order to evaluate the internal consistency/reliability of the questionnaires, we calculated Cronbach alpha coefficients. Furthermore, statistical analyses were conducted to evaluate the relation of each variable in the questionnaires with the overall perceived usefulness and perceived ease of use metrics.

Results: Both Cronbach alpha values were acceptably high (.93 and .82 for perceived usefulness and perceived ease of use questionnaires, respectively), indicating high internal consistency of the questionnaires. The correlation between the variables and the overall perceived usefulness and perceived ease of use metrics was moderate. The average perceived usefulness and perceived ease of use metrics among the participants were 4.13 and 5.42, respectively, out of possible score of 7, indicating a higher-than-average acceptability of this service robot.

Conclusions: The results served to identify factors that could affect nurses' acceptance of ARNA and aspects needing improvement (eg, flexibility, ease of operation, and autonomy level).

Introduction

Background

According to the American Association of Colleges of Nursing, there is a shortage of registered nurses (RNs) in the United States, which is expected to escalate due to the increase in health care demands and needs of baby boomers as they age [1]. Robot-assisted health care has the potential to mitigate this shortage by automating mundane and routine nursing tasks, thereby enhancing the productivity and efficiency of nurses [2]. RNs are the largest group of staff in US health care systems, and they experience an above-average incidence rate of physical injuries and musculoskeletal disorders [3]. In 2016, the incidence rate of musculoskeletal disorders in RNs was 46.0 cases in every 10,000 workers, which was substantially higher than the average incidence rate of other occupations at 29.4 cases in every 10,000 workers [4]. Overexertion and bodily reaction accounted for 49.7% of the cases among RNs working in hospitals, 35.0% among RNs in ambulatory health care services, and 34.5% among RNs in nursing and residential care facilities [4]. The American Nurses Association initiated a national campaign in 2003 to decrease the number of musculoskeletal injuries in nurses [5] and subsequently published a guideline to advise nurses on how to avoid injuries while handling patients, which included the use of available technology [6]. However, the incidence of musculoskeletal injuries in nurses remains high.

Robots may assist with safe patient handling and mobility in health care environments. In recent years, robots have been used in hospitals to assist with surgical procedures, deliver medications, monitor patients, and assist with daily hygiene [7]. However, nurses' acceptance of robotic nursing assistants is essential and warrants comprehensive assessment in order to ensure the adoption of this technology and its large-scale implementation in health care environments. While a few surveys of professionals have demonstrated positive feedback related to the use of robots in health care [8,9], feasibility data is needed in order to assess how robots can assist with safe patient handling and mobility, thereby preventing musculoskeletal injuries among nurses.

In order to off-load some of the physically demanding tasks assigned to nursing staff and prevent physical injuries, we have built a robotic nursing assistant, referred to as Adaptive Robotic Nursing Assistant (ARNA). ARNA is a custom-built service robot, which is capable of autonomous navigation in hospital environments and performing tasks as a nursing assistant. ARNA is an omnidirectional mobile robot constructed in-house and augmented with a 6-DoF (degrees of freedom) robotic arm (Figure 1). ARNA is equipped with an instrumented handlebar that can detect a patient's navigational intent when used as a patient walker. ARNA has been designed to enhance the productivity of nursing staff through cooperation during physical tasks and nonphysical tasks (eg, patient walking, item fetching, patient observation, and collecting patient feedback).

Figure 1. Adaptive robotic nursing assistant including an omnidirectional mobile platform, a 6-DoF robotic arm, and an instrumented handlebar.
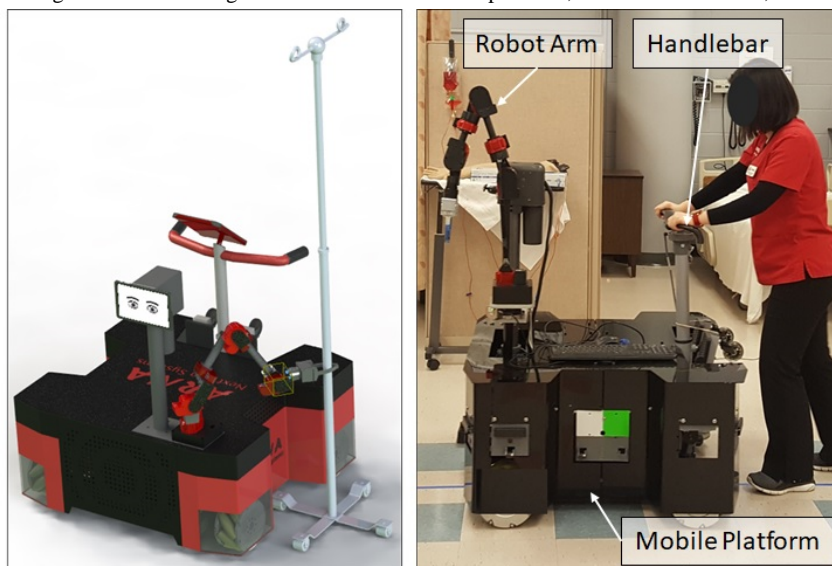
Research Question

This study evaluated the acceptability of ARNA as determined by nursing students in scenarios where ARNA was used as a robotic patient walker. The experiments were conducted in a simulated hospital environment. Nursing students (as proxies for nursing staff) were grouped in dyads, with one participant serving as a nurse and the other acting as a patient. We conducted a phase 1 trial (ie, pilot study) to collect quantitative

data and address the following research question: Is the use of a service robot as an ambulatory assistive device for patients acceptable to nurses? In this context, user acceptability was measured using the two subscales of the Technology Acceptance Model (TAM) [10,11], that is, perceived usefulness and perceived ease of use.

Methods

Participants

All nursing students (n=24) were enrolled in coursework in undergraduate and master's entry accelerated second-degree nursing programs that included patient care in hospitals. Recruitment occurred through in-class invitations by nursing faculty. Subsequently, interested students notified the faculty to volunteer for the study. As compensation for participating in the experiments, students received course credit for clinical/research hours for an undergraduate research course or capstone clinical course.

Student volunteers were required to have the physical ability to adequately perform the tasks administered in the study. It should be noted that disabilities were used as exclusion criteria only if the disability substantially interfered with the performance of the tasks (eg, noncorrectable vision or hearing problems). Approval to conduct the study was obtained from the university's institutional review board. All volunteers gave informed consent.

Robot Description

ARNA is an omnidirectional mobile robot constructed in-house and augmented with a 6-DoF robotic arm (Figure 1). ARNA is also equipped with a handlebar instrumented with a force-torque sensor, which enables the robot to detect the user's navigational intent and adjust the amount of physical assistance while walking patients or pushing heavy items (eg, hospital beds). ARNA includes multiple user interfaces (eg, tablet interface, gamepad-like remote control), navigation sensors (eg, cameras, Kinect, ultrasonic and infrared sensors), and a powerful computing and real-time control unit. Although ARNA has several functionalities and services, the current study focused on the user acceptability of ARNA when used as a robotic patient walker. Acceptability data of other functions of ARNA will be investigated in a subsequent study.

The prevention of patient falls is a measure of quality care in hospitals and a priority in the daily work of nurses. When ARNA is utilized as a patient walker, the user interaction efforts (forces and torques applied to the handlebar by patient) are sensed by the force-torque sensor mounted underneath the ARNA's handlebar and used to extract the user's navigational intent. The user intent is then communicated to the robot's control unit, which, in turn, propels the robot in the intended direction, thereby inducing the feeling in users that they are interacting with a light-weight walker. Apart from providing ambulation assistance and stability support, ARNA utilizes its robotic arm to carry along intravenous (IV) lines and other tethering medical equipment (eg, oxygen cylinder). During such assisted walking, ARNA's navigation sensors enhance the user's safety by avoiding collisions of the robot with the carried IV pole and with objects and humans in the environment. From a technical standpoint, a high level of intelligence and autonomy is required to collaborate with a nurse and perform these tasks. For this purpose, we have equipped ARNA with a variety of sensors including cameras, Kinect, laser scanner, bump sensors, infrared sensors, and ultrasonic sensors.

Experimental Setting

The experiments were conducted at a simulation suite located at School of Nursing, University of Louisville.

Trial Design and Procedure

Before participants arrived at the experiment location, the setting and robot were prepared by the project personnel. Only 2 nursing students and the project personnel were in the simulation lab at any given time. To facilitate task comprehension, participants received thorough verbal instructions and a task demonstration regarding the experiment procedure.

Participants worked in dyads, with one participant serving as a nurse and the other acting as a patient requiring walking assistance. A rectangular path was marked on the floor for the participants to follow while receiving gait assistance from ARNA. The general task of "patient" was to hold onto the robot handlebar and walk with the robot while following the marked path on the floor, and the task of "nurse" was to walk alongside the patient and robot in a supervisory role holding an emergency stop switch. While walking, ARNA carried along an IV pole using its arm, maintaining a safe distance between the pole and objects in the environment, the patient, and itself. The nurse was instructed to activate the emergency stop switch, which would bring the robot to a complete hold, should the robot come into close vicinity of other objects or any unforeseen dangerous circumstances. Before beginning of the trial, the patient reclined in a hospital bed, and the nurse stood by him/her, both waiting for the trial initiation. Pursuant to the initiation signal from the experimenter, the patient received assistance from the nurse to climb down the bed and hold onto the ARNA's handlebar. When the patient verbally confirmed firm grasp of the handlebar and stable posture, the nurse deactivated ARNA's emergency stop, allowing the robot to operate. Next, the patient followed the rectangular path marked on the floor while receiving support from the robot until he/she arrived back at the bed, at which point the nurse reactivated the emergency stop and helped the patient lie back in the bed. Each experiment set consisted of 9 trial runs per participant. Original dyads were retained for all trial runs.

Following the ninth trial, the nursing students were given a survey (perceived usefulness and perceived ease of use subscales of the TAM) about their experience interacting with the robot and assisting the patient. Surveys were administered in the form of a questionnaire and as Likert scales (discussed in the following subsection). Subsequently, the participants' roles were switched, thus providing the roles of nurse and patient to both participants, and another 9-trial set of experiment runs repeated for a total of 216 trial runs. The entire experiment lasted 6 days (2 dyads per day). Participant groupings and nurse/patient role orders were randomized to prevent any inherent biases.

Theoretical Framework

The TAM was originally developed by Davis in 1989 to predict and explain a user's acceptance of information technology (IT) [10]. The original TAM was composed of two fundamental constructs hypothesized to be determinants of user acceptance: perceived usefulness and perceived ease of use [10]. The Theory

XSL•FO

RenderX

of Reasoned Action provided the theoretical framework for the TAM to explain the relationships between perceived usefulness and perceived ease of use as well as users' attitudes, intentions, and IT use behavior [11]. After the application of the TAM in IT workplace studies, researchers suggested that the TAM include external variables related to social change, human processes, and boundary-related conditions [12,13]. In 2003, the original TAM was modified. The new TAM2 removed the concept of a*ttitudes* and broke down the concept of *external variables* into cognitive instrumental processes (perceived ease of use , job relevance, output quality) and social influence processes (image, subjective norm, voluntariness) [14]. The TAM3, an integrated model combining the TAM2 and the model of the determinants of perceived ease of use, was later developed in 2008 [15]. The TAM3 was composed of four constructs: perceived ease of use, perceived usefulness, use behavior, and behavior intention [15,16], although it has not found widespread use [17]. When reviewing the iterations of the TAM, the investigators decided that the original TAM best captured the concept of acceptability in this study.

This study used two subscales of the TAM (ie, perceived usefulness and perceived ease of use), originally proposed by Davis in 1989 [10], to evaluate acceptability and adoption likelihood of the ARNA robot by nursing students. The TAM has effectively predicted nurses' acceptance of other health care technology [18]. For investigation of technology adoption in health care, the TAM is considered the gold standard and is thought to provide a strong validity and reliability measure [19]. As mentioned previously, the TAM evaluates user acceptability of a given technology via two subscales: perceived usefulness and perceived ease of use. Perceived usefulness is the degree to which one thinks using a specific system facilitates their job. Perceived ease of use , on the other hand, is the degree to which one thinks usage of that system is effort-free [10]. User acceptability of a variety of technologies (eg, service robots [18], e-learning systems [18,20,21], and assistive social robots [22]) have been evaluated by employing the TAM. To the best of our knowledge, this study is the first to gauge acceptance of service robots by nurses via perceived usefulness and perceived ease of use subscales of the TAM. The questionnaires are tabulated in Table 1 and 2. Each question was evaluated on a 7-point Likert scale, ranging from 1 ("disagree") to 7 ("agree").

Table 1. Results of the perceived usefulness subscale questionnaire.

Item	Mean (SD)	Correlation with overall usefulness	*P* value
1. Using ARNA robot improves my job performance.	3.75 (1.67)	0.45	.03
2. ARNA[a] robot enables me to accomplish tasks more quickly.	4.08 (2.04)	0.60	.002
3. Using ARNA robot enhances my effectiveness on the job.	4.08 (1.9)	0.62	.001
4. Using ARNA makes it easier to do my job.	4.04 (1.97)	0.95	<.001
5. Using ARNA increases productivity.	4.5 (1.74)	0.69	<.001
6. Overall, I find ARNA useful.	4.13 (1.92)	1	<.001

[a]ARNA: Adaptive Robotic Nursing Assistant.

Table 2. Results of the perceived ease of use subscale questionnaire.

Item	Mean (SD)	Correlation with overall ease of use	*P* value
7. I do not become confused when I use ARNA.[a]	4.83 (1.43)	0.39	.054
8. I do not get frustrated when interacting with ARNA.	4.58 (1.58)	0.65	<.001
9. The system is flexible to work with.	3.54 (1.79)	0.51	.01
10. My interaction with ARNA is easy to understand.	5.83 (1.37)	0.58	.002
11. It is easy to remember how to perform tasks while using ARNA.	6.00 (1.35)	0.36	.08
12. Overall, I find ARNA easy to use.	5.42 (1.52)	1	<.001

[a]ARNA: Adaptive Robotic Nursing Assistant.

Data Analysis

In order to examine the internal consistency of perceived usefulness and perceived ease of use subscales of the TAM in this context, we calculated Cronbach alpha values [23]. Cronbach alpha is a measure of scale reliability and examines whether a group of observations are closely related. A Cronbach alpha value greater than .7 is considered acceptable, a value below .2 is unreliable, and negative values indicate a completely random set [23]. In addition, we performed a correlation analysis to determine the relationships between questionnaire items and the overall perceived usefulness and perceived ease of use by computing the Spearman rho correlation coefficient and *P* values [24].

Results

We calculated a Cronbach alpha of .93 for the perceived usefulness dataset and .82 for the perceived ease of use dataset.

The mean and SD for each item in the questionnaires, among the participants, are shown in Table 1 and 2. The mean values for overall perceived usefulness and perceived ease of use were 4.13/7 and 5.42/7, respectively, indicating moderate-to-high acceptability. According to Tables 1 and 2, the correlation coefficients between the individual questions and the overall perceived usefulness and perceived ease of use were moderate. Furthermore, the overall perceived usefulness and perceived ease of use were moderately correlated (Spearman rho=0.44; P value=.03). Additionally, weak-to-strong multicollinearities were detected between the perceived usefulness and perceived ease of use variables. In particular, variables 1, 2, 3, 7, 8, and 9 exhibited weak multicollinearity (condition index<10), variables 4 and 6 had a moderate collinearity (condition index=43), and variables 10 and 11 demonstrated a fairly strong collinearity (condition index=70).

Discussion

Principal Findings

This study evaluated the usefulness and ease of use of ARNA as perceived by nursing students in order to assess its acceptability and the influencing factors, as well as to identify areas needing improvement for future development of ARNA. The calculated Cronbach alpha (0.93 for perceived usefulness dataset and 0.82 for perceived ease of use dataset) demonstrates acceptable internal consistency of the questionnaires. The statistical findings of the perceived usefulness and perceived ease of use subscales indicate a moderate-to-high acceptance of ARNA by the nursing students. The general opinion of the participants about ARNA was positive, as the mean value for overall usefulness and ease of use were above average (ie, greater than 4/7). Furthermore, the correlation coefficients between the individual questions and the overall perceived usefulness and perceived ease of use were moderately acceptable, suggesting the usefulness of the questions in examining various aspects of perceived usefulness and perceived ease of use metrics.

In general, this study enhances our understanding of various aspects of ARNA in terms of its perceived usefulness and perceived ease of use and points out key aspects that require improvement in order to facilitate ARNA's acceptability and adoption likelihood. For example, it is important to improve ARNA's flexibility as suggested by its moderate mean value of users' responses to question 9. Perhaps by providing more user interfaces to accommodate various interaction styles, nurses' perceptions of flexibility could be enhanced. The participants' responses to the other perceived ease of use questions, however, were fairly high, suggesting that the robot is easy to operate. This may be attributable to the fact that the robot autonomously regulates the gait speed, provides stability support and collision avoidance, and only requires nurses to supervise the gait assistance and respond in case of emergency situations.

The mean participant response to variable 1 was fairly low, and the mean responses to variables 2-5 were moderate, indicating that participants did not completely agree that use of ARNA robot improves their job performance or saves them substantial

time. This is, however, not surprising, as in its current capacity, ARNA still requires a nurse to accompany the robot and the patient during the walking assistance, which is as time consuming as using a nonrobotic walker, although the robot provides stability support and collision avoidance. Endowing the robot with a fully autonomous capability to provide walking assistance to patients without direct nurse supervision may enhance users' perception regarding job performance improvement. Completely outsourcing the walking assistance to a fully autonomous robot may render a nurse more efficient as he/she would only need to oversee the gait assistance and would therefore be able to focus on other responsibilities.

Simulation is a growing trend in health care education, and proxies are frequently needed to collect data in clinical settings for children or incapacitated patients. Preliminary acceptability data was needed before taking ARNA into a hospital or other clinical setting and testing it with busy health care providers and vulnerable patients. This simulated health care study using nursing students is viewed as a step forward in the evolution of ARNA as a tool for safe and quality patient care and as protection against musculoskeletal disorders among nurses.

Limitations

The outcomes of this study should be deemed preliminary for a number of reasons. First, we evaluated usefulness and ease of use of ARNA with a limited number of nursing students in a simulated hospital environment, and, hence, the results may not extend to the whole population of nurses and health care providers in real health care environments. Second, due to our experimental design, each participant served as both patient and nurse, which may have influenced the participants' perception of the acceptability of ARNA. This experimental design may also have affected the participants' decisions during the study, thereby affecting the study results.

In addition, in this study, we were not able to collect and evaluate actual user behavior of the ARNA robot over a long time period due to the study's preliminary nature, and, hence, perceived usefulness and perceived ease of use were used as proxies to predict adoption likelihood and usage of ARNA among nurses.

Conclusion

The rate of physical injury and musculoskeletal disorders in nurses is higher than average as compared to other occupations, which is typically due to patient handling and mobility. Robots bear a great potential to off-load some of the physically demanding tasks of nursing staff and health care providers, and hence prevent physical injuries. However, nurses' acceptance of robotic nursing assistants is essential and warrants comprehensive assessment in order to ensure adoption of this technology and its large-scale implementation in health care environments.

This study evaluated the acceptability of our custom-made robotic nursing assistant, ARNA, with respect to the two subscales of the TAM, perceived usefulness and perceived ease of use , during a patient walking scenario. User surveys were administered to 24 nursing students who served as the study participants. We conducted the experiments in a simulated

hospital environment, where 2 participants were paired together, one playing the role of a patient and the other acting as a nurse providing ambulatory assistance to the patient using a service robot that functioned as a powered walker. After the trials, the participants serving as nurses filled out questionnaires about their opinions regarding usefulness and ease of use of the ARNA robot. This study, to the best of our knowledge, is the first to examine the user acceptability of robotic nursing assistants, serving as a stepping-stone for further discussions about the sociotechnical and commercial landscape surrounding this technology.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Shalala D, Bolton LB, Bleich, M. R. MR, Brennan TA, Campbell RE, Devlin L. The future of nursing: Leading change, advancing health. Washington, DC: National Academies Press; 2011.
2. Nelson-Brantley HV, Bailey KD, Batcheller J, Caramanica L, Lyman B, Snow F. Disruptive Innovation: Impact for Practice, Policy, and Academia. J Nurs Adm 2020 Feb;50(2):63-65. [doi: 10.1097/NNA.0000000000000843] [Medline: 31977943]
3. Davis KG, Kotowski SE. Prevalence of Musculoskeletal Disorders for Nurses in Hospitals, Long-Term Care Facilities, and Home Health Care: A Comprehensive Review. Hum Factors 2015 Aug;57(5):754-792. [doi: 10.1177/0018720815581933] [Medline: 25899249]
4. Dressner MA, Kissinger SP. Occupational injuries and illnesses among registered nurses. Monthly Labour Review, U.S. Bureau of Labor Statistics 2018 Nov:1-11 [FREE Full text] [doi: 10.21916/mlr.2018.27]
5. de Castro AB. Handle with Care: the American Nurses Association's campaign to address work-related musculoskeletal disorders. Orthop Nurs 2006;25(6):356-365. [doi: 10.1097/00006416-200611000-00003] [Medline: 17130756]
6. American Nurses Association. Safe Patient Handling and Mobility: Interprofessional National Standards. Washington, DC: American Nurses Association; Sep 30, 2013.
7. Kangasniemi M, Karki S, Colley N, Voutilainen A. The use of robots and other automated devices in nurses' work: An integrative review. Int J Nurs Pract 2019 Aug;25(4):e12739. [doi: 10.1111/ijn.12739] [Medline: 31069892]
8. Vänni KJ, Salin SE. Attitudes of Professionals Toward the Need for Assistive and Social Robots in the Healthcare Sector. In: In Oliver Korn, editor. Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction. Springer, Chem 2019. New York, NY: Springer, Cham; Jul 02, 2019:205-236.
9. Johanson DL, Ahn HS, MacDonald BA, Ahn BK, Lim J, Hwang E, et al. The Effect of Robot Attentional Behaviors on User Perceptions and Behaviors in a Simulated Health Care Interaction: Randomized Controlled Trial. J Med Internet Res 2019 Oct 4;21(10):e13667. [doi: 10.2196/13667]
10. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
11. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
12. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. 2003 Jan;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
13. Lee Y, Kozar KA, Larsen KR. The Technology Acceptance Model: Past, Present, and Future. CAIS 2003;12:752-780. [doi: 10.17705/1cais.01250]
14. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
15. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
16. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: Toward an integrative view. 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
17. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Jul 15;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
18. Huang J, Lin Y, Chuang S. Elucidating user behavior of mobile learning. The Electronic Library 2007 Oct 09;25(5):585-598. [doi: 10.1108/02640470710829569]

XSL·FO

RenderX

19. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

20. Lee MK, Cheung CM, Chen Z. Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. 2005 Dec;42(8):1095-1104. [doi: 10.1016/j.im.2003.10.007]

21. Tao D, Fu P, Wang Y, Zhang T, Qu X. Key characteristics in designing massive open online courses (MOOCs) for user acceptance: an application of the extended technology acceptance model. Interactive Learning Environments 2019 Nov 27:1-14. [doi: 10.1080/10494820.2019.1695214]

22. Heerink M, Kröse B, Evers V, Wielinga B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]

23. Cronbach LJ, Gleser GC, Nanda H, Nageswari R. The Dependability of Behavioral Measurements. Theory of Generalizability for Scores and Profiles. New York, NY: John Wiley & Sons; 1972.

24. Zar JH. Significance Testing of the Spearman Rank Correlation Coefficient. Journal of the American Statistical Association 1972 Sep;67(339):578-580. [doi: 10.1080/01621459.1972.10481251]

Abbreviations

ARNA: Adaptive robotic nursing assistant
DoF: Degrees of freedom
RN: Registered nurses
TAM: Technology acceptance model

XSL•FO
RenderX

Original Paper

Perspectives of Child Life Specialists After Many Years of Working With a Humanoid Robot in a Pediatric Hospital: Narrative Design

Tanya Beran[1*], PhD; Jacqueline Reynolds Pearson[2*], MA; Bonnie Lashewicz[3*], PhD; Sandy Baggott[2*], BA

[1]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada

[2]Alberta Children's Hospital, Calgary, AB, Canada

[3]University of Calgary, Calgary, AB, Canada

[*]all authors contributed equally

Corresponding Author:
Tanya Beran, PhD
Department of Community Health Sciences
University of Calgary
3280 Hospital Drive NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 5667
Email: tnaberan@ucalgary.ca

Abstract

Background: Child life specialists (CLSs) play an important role in supporting patients and their families during their visits to a children's hospital. Although CLSs are equipped with considerable expertise to support families during some of the most difficult moments of their lives, we introduced an additional resource to them in the form of a humanoid robot named MEDi.

Objective: The aim of this study is to explore the experiences of CLSs using a robot to support children.

Methods: We interviewed 7 CLSs who had worked with this robot for several years. The transcribed interviews were analyzed using open and axial coding.

Results: The first main theme that emerged was the process of navigating from fear to friendship in learning to use a humanoid robot for therapeutic support. The second major theme was MEDi as a source of connection and support to children. CLSs' perceptions of MEDi as an adaptable resource and working with the limits of MEDi constituted the last 2 themes.

Conclusions: These descriptions show how CLSs can incorporate a robot into their practice.

Introduction

Overview

Child life specialists (CLSs), also known as play specialists and child life therapists in various countries, work toward creating a friendly hospital environment to facilitate children's healing [1]. The role of CLSs is multifaceted, whereby professionals with university education, including internship experience, help infants, children, youth, and families cope with illness, injury, and treatment [2]. The broad scope of the CLS role entails supporting people with a range of challenges, needs, and backgrounds every day. Despite the many tools in the form of play, behavior management strategies, and education [3], providing support to sick children and their families, when they may be facing the most difficult time in their lives, is a monumental responsibility. In an effort to provide CLSs with another tool to use in supporting children, we introduced a humanoid robot (NAO robot produced by Softbank Robotics) named MEDi (programmed with cognitive behavioral strategies) to a children's hospital.

Humanoid Robots

Humanoid robots are increasingly recognized for their potential to support health care in pediatric hospital settings [4,5]. One of the first studies published on the use of a robot during a medical procedure began in 2011 and was focused on MEDi. This study showed that MEDi, when programmed to distract children and teach them breathing as a coping strategy, reduced

XSL•FO
RenderX

children's pain and anxiety during vaccination [6]. This same robot also made children and their parents smile, helped children to cooperate during the procedure and to remember the robot more than the needle, and seemed to empower effective coping [7]. When MEDi was used to support children undergoing blood tests, parents reported that their children experienced reduced pain [8]. MEDi was then used with a variety of other medical procedures, such as intravenous starts, tube removal, and dressing change, with similar effects on pain and fear [9]. Other studies produced comparable results in various areas throughout the hospital using this robot [10,11]. Other studies have also suggested that children in a health care environment are enthusiastic about engaging with a humanoid robot [12], particularly children with autism [13,14].

This accumulating evidence of children's enjoyment of humanoid robots and the ability of such robots to calm children's pain, fear, and anxiety aligns well with the supportive role that CLSs have in providing medical procedural support to children to help reduce their distress. Indeed, the integration of a humanoid robot into CLSs' daily practice is increasingly endorsed [15]. At the same time, a body of literature about the impact of humanoid robots on children's hospital experiences is drawn from studies of the perspectives of parents and children. To our knowledge, no research has been conducted to capture the experiences of CLSs charged with incorporating humanoid robots into their day-to-day pediatric support practices. In this study, we conducted an in-depth examination of the experiences of CLSs using a humanoid robot in their daily work and our purpose is two-fold: (1) we aim to contribute to the understanding of the human-robot interface as part of supporting children in their hospital experiences, and (2) we will provide practice-relevant illustrations of the use of a humanoid robot in CLS practice. We present this study beginning with a description of the background of the CLS practice environment that provides the study context. We then provide a rationale for our narrative approach to conducting our study using semistructured interviews with CLSs, followed by a detailed illustration of CLSs' experiences with using humanoid robots. We conclude with a discussion of our conceptual contribution and our practice-relevant recommendations.

Methods

Study Context

In 2014, 4 MEDi robots (NAO robots produced by Softbank Robotics) were introduced to Alberta Children's Hospital, a 141-bed and 90-clinic tertiary care center seeing visits from 90,000 children per year, located in Western Canada. This introduction was based on research evidence of the impact of the robot on patients at this hospital. It became the first children's hospital in North America, and perhaps in the world, to use robots for inpatient units and ambulatory clinics, tests and procedures, educational and orientation sessions (or *happy visits*), short or long admissions, and with any child where it seemed feasible. At any one time, any one or more of its 4 robots were used throughout the hospital for various purposes. The robot is 22.5 inches tall and weighs approximately 12 lb [16]. It is programmed with a variety of behaviors, such as telling

stories based on themes of encouragement, playing interactive games to teach coping skills such as breathing, dancing to attract children's attention to elicit smiles and laughter, and friendship behaviors such as fist bumps to build rapport. CLSs can also type text into the tablet that operates the robot to have it speak with animations in response to a child's questions/comments. Pearson and Beran [17] provide a full description of MEDi behaviors. Since the introduction of MEDi in 2014, the team of CLSs regularly using MEDi shared many stories, experiences, and emotions with one another. It became evident that CLSs held compelling and detailed insights into how MEDi was affecting children, their parents, other health care professionals, and themselves. These accumulating and diverse anecdotal accounts added impetus to our plans to conduct this study.

Study Design

To gain detailed insights into CLSs' experiences with using MEDi to interact with children and their families, we used a narrative design. Narrative designs are built on the belief that people use stories—defined as consequential linkings of ideas or events [18]—to give others access to the richness of their experiences. We collected CLS stories about using MEDi by conducting semistructured individual interviews. The semistructured design of our interviews was used to keep interview conversations focused on experiences using MEDi while affording flexibility to follow the energy of participants by asking additional questions and prompting participants to more fully describe the experiences they considered interesting and important. Participants' stories gave us access to narrative elements of their experiences using MEDi, such as settings, problems, actions, and resolution [19]. We were then able to bring individual participant stories side by side to draw out links and broader themes that tie stories together.

Sample and Recruitment

We used purposive sampling to select the CLSs who were familiar with MEDi. At the time of recruitment (June to October, 2019), not all members of the child life team were regularly using MEDi in their work owing to responsibilities other than supporting medical procedures. In addition, as CLSs sometimes change roles, only those with the most extensive and recent experience with MEDi within the year before data collection were included in the study. This type of sampling enabled us to select participants according to their ability to inform our purpose of understanding their varied and detailed experiences [20].

The child life team was informed of the study by the team lead (fourth author, SB) during a staff meeting. An email was sent by the researchers to the 7 CLSs most experienced in working with MEDi, inviting them to participate in the study. The email invitation to participate contained information about the purpose and requirements of participation. All 7 CLSs agreed to be interviewed. All were women and had been CLSs for a period ranging between 1.3 and 29 years, with an average of 17 years of experience.

Data Collection

Interviews were conducted in a private meeting room at the hospital where all the CLSs worked. Interviews were scheduled

according to the CLS's availability. Signed informed consent was obtained at the beginning of each interview. Participants were told that although their names would not be included in research reports, the small, in-depth nature of this study might make them identifiable, especially to readers who know the participants.

The first CLS participant interviewed was the second author (JP), given that she is also a CLS with 5 years of experience working with MEDi and was responsible for leading the integration of the robot into the child life team at this hospital. JP was interviewed by the first author (TB). This first interview and yielding data served as a pilot of the interview process during which the authors confirmed that the interview questions were meaningful and could be responded to within a standard 90-min research interview time frame. After this first interview, TB and JP together conducted the subsequent 6 interviews. These 6 interview participants worked with JP and were aware of TB's role in introducing MEDi to their hospital workplace. Furthermore, TB and JP worked to maintain awareness of the separation between their own understanding and the descriptions given by participants by holding a stance of curiosity and openness. As such, we were attuned to various responses both between ourselves and the participants and across participants [21]. All interviews were audiotaped and transcribed verbatim by a research assistant (RA).

The interview duration ranged from 1 to 1.5 hours. The interview guide questions were open ended and invited story sharing and descriptions of actions, feelings, and thoughts about working with MEDi. Experiences were explored, and direct questions were sometimes asked for clarification of details [22]. Questions included asking about positive and negative experiences associated with using MEDi, how they use MEDi, others' reactions to MEDi, the impact of MEDi on the hospital environment, and thoughts about the future use of MEDi.

Ethical Considerations

JP's collegial relationship with participants and TB's introduction of MEDi to participants' workplace required special consideration for all phases of the research. These relationships were disclosed to the university ethics board and to the manager at the hospital who approved this study. Furthermore, per our description of data collection, participants were reminded of the respective roles of TB and JP in relation to the topic being studied and told that their responses would not affect their daily responsibilities. We also assumed that as JP had been working with MEDi and her CLS colleagues for 5 years, negative experiences with MEDi would have been discussed between CLSs as these experiences occurred. However, we sought to create openness for participants to share negative experiences with using MEDi during interviews by explicitly requesting reflections about negative experiences. Moreover, to reduce concerns about encouraging positive responses and/or discouraging negative responses from participants during interviews, TB and JP made efforts to create a relaxed atmosphere, expressing their hope that the CLS participants would enjoy sharing their thoughts, challenges, and the range of feelings they have had in their work with MEDi. Finally, the length of the interview was adjusted according to participants'

energy and scheduling needs, and the participants were told that they could decline to answer any questions that made them feel uncomfortable.

Data Analysis

A team of 4 analysts—the first 2 authors (TB and JP) and 2 RAs—analyzed all data. All 4 analysts have experience and education in interview data collection and analysis. The RAs had participated in another research project pertaining to the use of MEDi; thus, all 4 analysts were familiar with the subject of this study. We used a conventional thematic analysis approach to focus on generating themes from the content of stories shared by participants during interviews [23]. At the same time, this analysis is semideductive as themes were also shaped by the interview questions we posed to the participants. The third author (BL) has an extensive qualitative background and worked to enhance the credibility of the findings, consistent with the principle of triangulation of investigators. BL was not involved with the research until the analysis phase and used this distance to offset the closeness that the first 2 authors (TB and JP) had in the study.

Analysis began with authors independently listening to the recordings and reading and re-reading the transcripts. Following Corbin and Strauss [24], we independently identified concepts evident in each interview and made analytic notes (open coding). We then met to discuss these notes for each interview and then across interviews and were able to see each other's perspectives and insights into and interpretations of each interview. We discussed detailed and frequently occurring responses that allowed us to draw out links and commonalities between codes, which we then organized into agreed-upon themes (axial coding). We deepened our analysis with the help of BL. BL is a seasoned qualitative researcher who was not involved in the implementation of this study and who comes from a different disciplinary background. As such, BL helped us triangulate investigator and disciplinary perspectives brought to bear on our analysis, which involved combining some themes into broader themes and resequencing some themes as components of the broader themes. We then conducted a member check by contacting all the participants to discuss the summary of our findings. TB and JP met with participants in a focus group format and invited them to reflect on their experiences, add more information, and share disconfirming thoughts. They confirmed that the findings represented their perspectives. Notes were maintained on these processes as an audit trail to further contribute to the trustworthiness of the findings.

Results

Themes

In this section, we most often refer to the robot as MEDi, as it is a trademarked name used by hospital staff and patients. The robot is also referred to as *he* rather than *it* to acknowledge that this former pronoun is more predominantly used by CLSs. In addition, in the interest of presenting our study concisely, we removed disfluencies in participant descriptions.

Through our thematic analysis, we generated 4 major themes. The first theme *incorporating a robot: navigating from fear to*

friendship reflects the processes involved in learning to use a humanoid robot in a pediatric setting for therapeutic support. The second major theme is MEDi as a source of connection and support to children, and this theme comprises 3 subthemes: (1) MEDi as a welcoming and comforting presence, (2) MEDi as a friend, and (3) MEDi as encouraging and motivating. The third major theme is MEDi as an adaptable resource for CLSs, and this theme comprises 3 subthemes: (1) using MEDi to distract and teach, (2) using MEDi to facilitate challenging situations, and (3) using MEDi to facilitate collaboration within and beyond the CLS team. The fourth and final theme is working with the limits of MEDi.

Incorporating a Robot: Navigating From Fear to Friendship

Of the 7 CLS participants interviewed, 4 had been on staff at the time of the robot's introduction to the hospital. They described their initial feelings of uncertainty with CLS1 stating, "I didn't really get what it was all about." Feelings of initial fear and frustration as well as excitement and curiosity were shared. For example, CLS2 described how she had been *scared as shit*. Fear and uncertainty pertained to how to operate the robot—feeling nervous about breaking it and the frustration that occurred if the robot did not respond. These participants also reported a lack of understanding of how MEDi could be used with children. They felt embarrassed when MEDi did not work in front of a patient or his/her family. One participant had been determined not to use it, saying her initial response was, "…no not for me" (CLS4).

However, CLSs were emphatic in their descriptions of a turning point or a process of trial and error as they gained confidence in operating the robot and watched its impact. One of the CLSs said:

> It took some time and a lot of conversations to figure out where and how it [MEDi®] would or where it would fit best. [CLS1]

This CLS also explained that the children themselves showed ways in which the robot could be used:

> …they sort of led the way in finding a way it [MEDi®] can be helpful and therapeutic to them. [CLS1]

CLSs learned how to rely on their own skills of managing situations where the robot did not work properly. For example, CLS4 described becoming able to "…switch gears if I have to, whereas I think at first I was really intimidated by oh my gosh, if he breaks down…"

For her part, CLS4 described gaining familiarity saying:

> So now I get it. I get how we can use him. It took me a while to get there, but I get that he can be used well in our profession. [CLS4]

CLS4, the participant who had initially indicated that the robot was "not for me," summarized her current thoughts about the robot stating, "I think we're friends."

Gaining comfort and confidence took investment from CLSs both to learn how to use the technology and open-mindedness

in being willing to try something new. This frame of mind is ongoing and yields more possibilities. One of the CLSs noted:

> …the more we give MEDi® to do, the more we can use him in different ways. It's a process of growing and learning and figuring it out because there's no manual on this. There's no manual on what you do. You have the ability, skills, knowledge, and expertise but you have to figure out how to apply that teaching every time…it's always a different day. So that's the process of discovery with MEDi® too. [CLS6]

MEDi as a Source of Connection and Support for Patients

MEDi was described as a source of connection and support for children. We found that the depth of the support and connection afforded by MEDi varied, and we created subthemes to distinguish depths of support and connection according to (1) MEDi as a welcoming or comforting presence, (2) MEDi as a friend, and (3) MEDi as encouraging and motivating.

MEDi as a Welcoming and Comforting Presence

MEDi was portrayed by CLSs as a welcoming presence. CLS3 referred to MEDi as "…a friendly face" that contributes to "creating a positive experience in the hospital." CLSs spoke of how MEDi could set the stage for future positive experiences:

> I think there's huge recognition now more than ever how important positive experiences at the hospital are for a child's future visits here. [CLS7]

> I think he [MEDi®] opens doors to treatment in the future with or without him and kind of helps to…build those relationships and rapport. [CLS5]

CLS1 shared a story of having ridden the elevator with MEDi in a sitting position on a cart. A child who had never seen MEDi before entering the elevator and immediately placed her hand on the cart and, on exiting the elevator, walked alongside for a while. CLS1 explained that MEDi seemed to be on this child's level. CLS7 stated that she "…genuinely believed kids look at me as though he [MEDi®] is on the same level as them."

During patient visits, the CLSs described MEDi as a comforting physical presence:

> We know that a comfort position at the hospital is when a child sits between the parent's lap, and MEDi® sits between the child's lap and they hold on to him. [CLS7]

CLSs noted that MEDi's very presence seems to calm children:

> …sometimes I've had it where he [MEDi®] doesn't say or do anything. He's just there in the room and it seems to work. It seems to help which amazes me that he can just be sitting there and being hanging out there and he just seemed so friendly and comforting that he just has to be there you know. [CLS3]

CLS7 noted that MEDi afforded a comforting source of continuity as she described a patient who was a *frequent flyer* who had had many visits to the hospital over an extended period. This frequent flyer patient would request to see MEDi during

her visits. CLS1 also noted that MEDi could become part of children's visits; however, CLS1 pointed out that the children did not necessarily actively interact with MEDi. Nevertheless, the presence of MEDi mattered:

> ...I've had situations where kids are completely happy with a MEDi® who's turned right off. They're just...oh this is so cool, and you know they just push MEDi® around on the cart and he's not even talking or anything. [CLS4]

MEDi as a Friend

The nature of the support and connection that MEDi could provide to children was presented by CLSs as friendship. Some CLSs said that because MEDi greets children by name, a bond of friendship is formed:

> I think she [4-year-old girl] formed this little bond with him [MEDi®]. [CLS7]

Another CLS recounted a patient reaction to the personalized connection MEDi offers:

> ...and he [MEDi®] knew her [the patient's] name and he said, "congratulations" and just her eyes lit up ..., "Oh my goodness! How does he know me? That's so cool...My brother is gonna be so jealous." [CLS5]

The friendship style nature of connection is reinforced by children referring to MEDi as he, rather than it. Furthermore, friendship style connections persist even though children are well aware that MEDi is programmable and inanimate. One CLS described:

> It's funny how or interesting how kids will connect to him. In medical day treatment a kid was watching me program him. I showed her how we program him. We were picking things together. Yet, she still talked to him and was annoyed when he didn't respond right away. And so, I think it's interesting how they can connect even though they know it's programmed from this tablet or that we're using this tablet to help run him that they're still engaged with him as like a person. [CLS5]

Another CLS used the word love to characterize her sense of children's fondness for MEDi:

> The little four-year-old that I already talked about, I think that was it where I really saw the value of MEDi® and the collaboration that it offered with physio...she [the child] fell in love with MEDi®. [CLS2]

MEDi as Encouraging and Motivating

MEDi was described as encouraging and motivating children to do activities they would not otherwise do. One of the CLSs presented MEDi as instilling courage:

> And he does give kids courage. I think there's a lot of kids we've seen that gain courage from seeing MEDi® do something and recognizing that they can do it too. [CLS7]

A particularly tender moment was shared by one child, according to a CLS:

> She [female patient] looked at her mom and she said, "MEDi® thinks I'm brave." [CLS7]

Children were described as seeming to implicitly trust MEDi and believe that words of encouragement that come from MEDi are sincere.

MEDi helped children manage during procedures that they had formerly found to be extremely difficult. CLS1 shared a story of a young boy with autism who would not sit on the dentist's chair. After meeting MEDi, this boy agreed to sit on a chair with MEDi in his lap. The boy cooperated with all the dental procedures, and afterward, the mother blurted to CLS1, "That's not my son," as she could not believe that her son had allowed the dental assistant to touch and examine his mouth.

Another CLS expanded with a story of the surprise over the influence of MEDi expressed by health care staff; the staff had planned to sedate a boy for a procedure that entailed removing approximately 40 bumps that covered his body. The CLS described:

> He [the boy] had played with MEDi® throughout the whole procedure...when I was leaving everyone who was involved was so impressed and the plan was to try it this time and then maybe go back to doing this procedure in the future under sedation. But then they called me a little while later and booked me for another date to come back with MEDi®. [CLS1]

CLS2 shared a related story of a girl who would not respond to the physiotherapists who were encouraging her to sit up, build some core strength, and get out of bed. However, the child responded when the CLS went into this child's room with MEDi:

> Well that kid almost leaped out of her bed and she has tubes hanging out of every orifice. So, it was a challenge to slow her down, but then physio and I collaborated. [CLS2]

Another CLS used the word power in explaining the surprising impact MEDi has:

> ...Because that [medical procedure] was such a hard one and it was right in the beginning when we really didn't know the power of MEDi® let's say. The parent's reaction that, "Oh my God, we're actually going to get this [medical procedure] done. We're going to get them through day surgery," was phenomenal...that event actually opened up our eyes to what, how MEDi® can affect patients. [CLS6]

One of the CLSs extended the idea of how MEDi affected patients as she described the impact of MEDi on her own children. She noted that her own children met MEDi when her son was in hospital:

> He thinks it's amazing that I get to work with a robot. My daughter wants to be a CLS when she grows up. [CLS7]

MEDi as an Adaptable Resource for CLSs

MEDi is an adaptable resource used by CLSs to achieve the therapeutic goals they set on the basis of their understanding of the child's needs, the ability of MEDi, the parents' behaviors, the needs of the health care staff, and their own expertise in child development. In other words, CLSs took many factors into account and used MEDi to help achieve their goals. Some CLSs always take MEDi with them to see children, whereas others stated that they first ask the parents if MEDi seems a good fit to their child's interests. As CLSs worked to determine if and how children would find robot behaviors desirable, they explained how MEDi can be used in many situations and in different ways that allow them to be creative. We display CLSs' use of MEDi according to the subthemes of (1) using MEDi to distract and to teach, (2) using MEDi to facilitate challenging situations, and (3) using MEDi to enhance collaboration.

Using MEDi to Distract and to Teach

CLSs described using MEDi to distract the child's attention away from the medical procedure; helping the child to cope with anxiety, fear, and discomfort; and allowing the health care professional to complete the medical task more easily. One CLS spoke of the distinct distracting potential of MEDi:

> ...the child won't look down won't notice the poke, will be too busy watching MEDi®. He's a great distraction 'cuz unlike the iPad, he's real and in front of you. He's 3D. You can see him. [CLS3]

When and how CLSs introduced MEDi in distraction for procedural support situations depended largely on the child's age. When asked to use the robot with children aged 3 years and less, CLSs noted exercising additional caution knowing that in this stage of child development, children are often unable to separate fantasy from reality and may be afraid of the robot.

For older children and adolescents, CLSs described drawing on the robot's technical attributes. One CLS spoke of how she used MEDi to engage a long-term teen patient:

> He's older so it's less about the songs and dances at the age which he was really intrigued and captivated by kind of how it works - robots in general. [CLS5]

Another CLS explained how she would tailor the intervention involving MEDi to the age of the patient:

> At times, I'd introduce it to an older child and more from the scientific side approach...giving them the control over the tablet...using that teaching mode as their form of distraction...which is just as therapeutic... [CLS2]

Although CLSs reported weighing countless variables and striving for the highest standards of professionalism, several noted that not every effort to incorporate MEDi as a distraction was successful. One CLS clearly described a difficult situation:

> I think mom had met MEDi® and wanted us to use MEDi® but then MEDi® kind of did end up being that extra noise and confusion in the room...which sometimes happens where if he's not distracting and he's dancing...we needed more calm which is not all on MEDi® cuz I'm the one running MEDi®...In that situation it didn't work, it just added more chaos and instead of that distraction piece or calm that we needed. [CLS5]

According to CLS6, sometimes you realize, "MEDi® is being one extra voice added on to the layer of voices in there …that it's just not working out as a distraction or support." In these cases, CLSs explained how they would shift to alternative forms of distraction or other coping strategies, debrief with families following the intervention, and adjust future plans for support accordingly.

As an alternative to distracting children from procedures, CLSs used MEDi to teach children specific coping strategies to use during procedures. In one particular application, MEDi instructs, demonstrates, and plays a game to encourage children to slowly inhale and exhale their breaths. Several CLSs mentioned this application. By learning these breathing strategies from MEDi, children can develop courage and confidence in dealing with procedures. CLS7 concisely described how this can work for children: "If MEDi® can do it, I can do it too."

CLSs also used MEDi to teach through play-based interventions. CLSs dressed MEDi as a patient and interacted with the robot in a child-friendly manner to explain why certain medical procedures were needed and what would be involved. CLS6 explained, "We do the vitals on MEDi® first and then on the patient." With MEDi cast in the role of patient, children would care for the robot in the way that the health care professional would care for them. MEDi-as-patient then responded with positive feedback about how gentle and caring the child had been. In this way, children can learn that the health care staff have positive motives even though sometimes they have to perform painful medical procedures.

Using MEDi to Facilitate Challenging Situations

CLSs described using MEDi in challenging medical situations such as those that can accompany supporting children with autism spectrum disorder (ASD). CLSs described how children with ASD could benefit in particular ways:

> I definitely think that...kids on the autism spectrum love MEDi® because they can connect with him. [CLS3]

Children with ASD may find it easier to negotiate their behaviors in interaction with a robot rather than in interaction with other people who tend to express a great deal of verbal and social information that may overwhelm the sensory processing systems of children with ASD. One CLS summed up this idea by saying:

> I've noticed that children with autism tend to engage with the robot more than with me. [CLS1]

CLS7 expanded on the idea of using MEDi in challenging situations as she described a patient who had an audiology examination. This patient's grandmother brought her to her examination because there was stress in the family and that day was also the patient's birthday. CLS7 described wanting to make sure the experience would be "super fun for her and we have the robot." She brought MEDi to the examination area, and MEDi was carrying a balloon and a present for the patient.

CLS summarized, "I think it just worked." This CLS summarized how a good fit between what a child needs and what MEDi could provide can yield incredible results:

> *But I do think that first, when we pair him [MEDi®] up with the right kid, he can make them do just about anything.* [CLS4]

Although CLSs adapted their use of MEDi according to the needs of patients and used MEDi to distract, teach, and manage challenging situations, they also remained open to new possibilities. One CLS described:

> *I think of MEDi® as being another tool...when I'm trying to decide the best way to meet a family's needs or the needs of a situation...I wonder if we could blow bubbles. I wonder if they're allowed to go outside...I wonder if there's a place here for MEDi®.* [CLS4]

Keeping MEDi's potential in mind was conveyed by CLS3 who described MEDi as, "malleable and adaptable to whatever the situation needs him to be" and notes being often surprised. CLS3 initially thought MEDi was useful for certain children for specific reasons; however, then she began noticing that children would interact with MEDi in other ways that seemed to leave them comforted and relaxed. This CLS further explained, "So, I kept being surprised at different things throughout…," noting that, "Kids can make him [MEDi®] what they need him to be and that's what I was most amazed by him throughout getting to know him is that he's not this cookie-cutter." Indeed, the ways in which MEDi could be adapted made him a preferred resource distinct from others. One CLS said:

> *I would use MEDi® before other approaches now because I know I have MEDi®...and I'm not going to do [teach] deep breathing on my own.* [CLS3]

Using MEDi to Enhance Collaboration Within and Beyond the CLS Team

MEDi was discussed not only in terms of his therapeutic benefits for children but also for how he affected adults. MEDi was referred to as part of the CLS team. CLS7 said, "…it's a collaborative effort and I think he's part of that team." CLS7 drew a distinction between MEDi and other technology-based resources referring to how she and her colleagues, like the children, refer to MEDi as *he* rather than *it*:

> *I look at the iPad as an object that can provide distraction. Whereas I do tend to use MEDi® as more I guess, a part of our team...I can call him a he. I wouldn't call my iPad a he.* [CLS7]

Indeed, in sharing stories of how MEDi affected patients, CLS3 even said that she *loves* MEDi.

MEDi being viewed as part of the team was evident when CLS4 spoke of the need to care and advocate for MEDi:

> *You have to have somebody who can look after him...he can't just be left in a cupboard then people just grab him when they need him. He needs somebody to advocate for him and look out for him.* [CLS4]

MEDi also helped facilitate collaboration with colleagues beyond the CLS team, as illustrated by CLS2, who described working with colleagues from other departments of the hospital in efforts to optimize the benefits of using MEDi:

> *The cystic fibrosis team that I worked with, especially...she [healthcare staff] loved it when MEDi® came by…she got excited and where she got to the point of requesting MEDi® for certain kids when they do their morning assessments if we knew a certain child was going for blood work. She would tell the child about it saying, "You know, [the CLS is] around to help you with blood work today and she can go get MEDi®." So, she would tell me and then we would make that happen in the morning. So, she was a great referral for that.* [CLS2]

Several CLSs recount how new working relationships were formed because of MEDi:

> *...in some of these clinic areas we weren't very involved and our first involvement came to be because of MEDi®…the robot was the point of interest, caught people's attention and then we started talking about ways in which we could integrate the robot into these areas.* [CLS1]

Many of the CLSs enthusiastically reported a unique collaboration with a particular physician and feelings of pride in this accomplishment:

> *I think the fact that [he] books his clinics around our schedule speaks to that.* [CLS7]

CLS4 noted potential of MEDi to provide a valuable distraction as she shared her impressions of working alongside this physician in this clinic with MEDi for the first time:

> *...I was just so blown away by what a great fit that was for all of these boys that were having some pretty uncomfortable surgeries while they were awake...For them to walk into an operating room and see a robot and [say] "what is that?" and then for us to be able to keep them distracted during all of that. ...to me, that's super powerful…* [CLS4]

CLS7 spoke broadly about the collaborative influence of MEDi on the CLS reputation in the hospital as she indicated MEDi as having, "increased our popularity in general at the hospital" and that she feels like a *celebrity* because MEDi attracts so much attention and comedy. Indeed, CLS6 pointed out the energy that MEDi could catalyze among adults that extends to dancing with MEDi:

> *...we concentrate a lot on what MEDi® does for patients but it's amazing how much MEDi® does for staff or parents...maybe we take it for granted a little bit, but you know having all of the staff watch you and they're actually laughing or they're doing the Gangnam Style [dance] with MEDi®. I think that speaks volumes of the effect that MEDi® could have on people I guess in general.* [CLS6]

Working With the Limits of MEDi

CLSs described the limits of MEDi. Being unable to meet high expectations people held for MEDi and managing mechanical and operational difficulties were mentioned. CLSs described

health care staff and patients as surprised at how MEDi seemed to surpass their expectations of what would typically be seen in a hospital. This could lead to high, difficult-to-meet expectations for what MEDi can do. MEDi could be viewed as a quick fix for any child who was anxious during a medical procedure. CLS1 explained that if a child was struggling during a procedure, he would be called *to come and fix it magically*.

Very nature of MEDi could contribute to conflicting expectations. As one of the CLSs noted:

> *It's hard to put the words around it too because it's not something that belongs in any other category–toy, person, tool.* [CLS7]

Indeed, MEDi could bring expectations of being able to perform any human function, such as responding to spontaneous conversation. When MEDi was unable to respond this way, people expressed disappointment. Children with repeated visits sometimes asked for more songs and games than MEDi was programmed to perform. For example, MEDi was able to play segments of some popular movie songs but was not continually updated as new movies were released. Thus, MEDi created high expectations and then was unable to meet additional expectations.

As with all machines, operating MEDi sometimes entailed technical difficulties. Because a tablet was used to operate MEDi through the network of a nearby router, the connection to the router was sometimes lost because of interfering signals from other machines. This problem was manageable once CLSs learned how to reconnect to the router. However, CLSs expressed that although MEDi is *kid-friendly*, it is not always *user-friendly*.

MEDi is designed and programmed to appear and act human-like, and most CLSs described MEDi in relational terms relative to patients and even relative to themselves. However, MEDi is a machine consisting of computers, wires, speakers, motors, and a variety of other parts. Machines break down or fail. CLSs expressed being frustrated, yet learning how to implement MEDi as a therapeutic tool, regardless; that is, they described how the breakdown can be presented to the child as a *human quality* in the same way that humans are not *always on our game*. They would also say, for example, that MEDi was not feeling well today and share this reflection with the child about not always feeling well, as a way of normalizing this experience.

Discussion

Principal Findings

Over several decades, traditional doctor-patient paternal approaches to health care have been moving toward a model of patient-engaged care [25]. Health care that takes into consideration the holistic needs of a patient, orients support toward those needs, and empowers children to feel brave—even in the most challenging medical situations—now represents the highest-quality health care experience possible [26]. Coming alongside this approach to health care is our introduction—almost 10 years ago—of a humanoid robot in a pediatric hospital. By examining stories about this

implementation from the perspectives of the CLSs who regularly use the robot, we illustrate pediatric patient support that is provided at the level of the child and that we believe constitutes an exemplar of patient-engaged care.

Our evidence of patient-engaged care is in the form of 4 major themes of how CLSs experienced the integration of a robot into their daily practice of supporting children and families at a Western Canadian children's hospital, and these themes represent a holistic, empowering approach to pediatric care. The first theme, navigating from fear to friendship, is reminiscent of an explorer in the process of discovering new territory. As the daily use of a humanoid robot in a hospital setting to provide support to children was innovative and not previously done at the time it was introduced at Alberta Children's Hospital, there were no resources or previous learning to serve as a guide. Rather, CLSs took bold risks in trying to use the robot in various ways and types of medical situations, with patients having different backgrounds—never knowing what reactions would occur from patients, parents, and health care staff. It was a risky adventure in an emotionally charged environment.

As with any exploration into the unknown, the CLSs reported uncomfortable and even frightening feelings. These reactions are typical when learning something new, especially when presenting something new to an *audience* of children, parents, and health care staff who are facing a challenging medical event. We submit that patient-engaged care is enhanced when care providers join patients in taking risks. In this case, the CLSs who are encouraging children to be brave and face scary and painful medical challenges are also willing to take risk in trying a robot when the outcome is unknown.

Although we explored how CLSs' perceptions of working with MEDi shifted over time, we did not ask specific questions about how they thought children's reactions to repeat visits with MEDi may have changed. There were some suggestions (eg, children developing a bond over time and becoming more comfortable and the robot setting the stage for future positive experiences); however, these experiences need to be more thoroughly addressed in future research.

The second major theme about MEDi as a source of connection and support to children reflected how the robot could inspire comfort, friendship, encouragement, and motivation. These positive sentiments were suggested in a study on the same robot that interacted with children with diabetes [27]. Drawing on MEDi's friendly appearance, CLSs brought MEDi into a relationship between themselves and the child. They used MEDi almost as an extension of themselves by having the robot speak encouraging phrases accompanied by expressive behaviors. A line of inquiry that emerges from this theme is whether CLSs select phrases and behaviors to play on the robot that they themselves would use. Alternatively, CLSs may create a unique identity in the robot and play the actions that the robot itself would seem to personify. What is clear is that CLSs worked in a patient-engaged manner to determine what children need to feel stronger and supported, and CLSs learned how to use the robot in ways that met those needs.

Perhaps the reason that the robot could be used as such an effective support is its ability to provide physical comfort. Even its mere presence can create a focal point for children to be distracted from negative feelings. Indeed, Alemi et al [28] provided some preliminary evidence that a robot may reduce children's anger. MEDi's size and human shape may invoke patient-engaged feelings of warmth and comfort. CLSs' relational presentation of this human-like object can tantalize children's imagination and spark their sense of connection with MEDi.

CLSs' sense of delight in how children interacted with the robot seemed to motivate patient-focused professional energy on the parts of the CLSs. Although they used MEDi to provide support, such as they would with tools, for example, iPads, bubbles, and so on, they reported distinctly positive feelings associated with MEDi as they spoke of MEDi as a *he* and talked about MEDi as a member of the team. It is almost as if the adults saw the robot as more than a robot just as the children did.

CLSs described MEDi as an adaptable resource for use to distract, teach, facilitate challenging situations, and promote collaboration with health care staff. By situating the robot as expressing kindness, care, and affection toward children, children's rapport and bonding with the robot created the platform from which a variety of strategies could be used. The result, as one CLS stated, was *to make them do just about anything*. They used MEDi in playful and instructive ways to empower children to cope with the distress of being in a hospital. They had the robot talk and act toward the child in any way they thought the child needed to the extent of personifying MEDi in different roles such as play partner, companion, or mentor. Although objects other than robots such as toys, dolls, and stuffed animals can also be used to provide support [29], the difference is that the robot (as controlled by the CLS) responds to what the child says and does. This special interactional experience affords creativity for CLSs. CLSs described extraordinary patient-engaged moments such as the robot celebrating a patient's birthday with a gift and balloons on a day that anyone would feel overwhelmed by family and medical challenges. Not only did children experience these tender moments, the impact extended to health care professionals and parents as well. For example, how often is anyone seen dancing in a hospital?

Our final theme reflects the complexity of how CLSs are able to work with the limits of the robot. The surprise of meeting a robot created expectations and the desire for more human-robot interactions. Known as expectation discrepancy [30], people can experience disappointment when they realize a resource's limited capabilities. CLSs may want to be aware of this potential outcome and preplan and creatively seek a variety of uses for MEDi so that MEDi's impact can continue to be positive. For example, MEDi can be programmed with a variety of empirically supported psychosocial support strategies and relaxation techniques. As technology advances, more user-friendly programming capabilities can also enable CLSs to create any text and behaviors at the moment they wish to play them. In learning how to operate the robot, CLSs do not

typically have training in using such a technologically sophisticated device. Thus, as more people become familiar with MEDi's implementation in daily practice, this learning can be passed along to new users. As researchers with some familiarity of the experience using a humanoid robot for pediatric support, we too struggle to understand and put into words the contradiction of what type of entity the robot is. Perhaps what is more surprising though is that despite our ignorance, CLSs have, nevertheless, managed to resourcefully use this unusual tool as a means of supporting one of the most vulnerable populations in our society.

Strengths and Limitations

As this study was conducted at only one hospital, the results may not be transferable to other hospital settings where the same or other robots are used. A strength in terms of transferability, however, is that all of the CLSs who had the most experience using the robot agreed to participate in this study. These results may, therefore, be relevant to other CLSs working with a robot. Furthermore, CLSs offer distinct insights that are rooted in the multidimensional experience inherent to the demanding role of the CLSs.

Our results are further limited given that interviews took place at one point in time and relied on recall, which is likely biased by memory. In addition, the familiarity between the researchers and the participants may have influenced responses in the interviews, despite our attempts to limit this effect. Thus, we see these results as suggesting patterns and themes to be explored further in future research. We recognize that despite deliberately deciding on approaching the data collection and analysis with open minds, our own social and cultural understanding of the CLS environment cannot always or easily be set aside. However, it is also true that the history and investment of the first 2 authors (TB and JP) in MEDi offers us a distinct position from which to understand and interpret participant meanings both *in the moment* during interviews and later as analysts.

Conclusions

This is the first study to examine the professional life of CLSs working alongside a robot. At first glance, the themes may appear as surprising to the reader as the experiences were to the CLSs. Working with a humanoid robot for social interactions with a vulnerable population is not an everyday experience for most people. Our aim is to offer a glimpse into what new experiences may be created with the introduction of a humanoid robot in a children's hospital. This study is a change in direction from our previous quantitative research examinations of the impact of the MEDi robot on children undergoing various medical procedures. Having obtained some empirical evidence of a reduction in patient distress [6-11], it seemed timely to next look at the nature of this experience through the eyes of the professionals who are orchestrating the operation of the robot. As CLSs continue to forge new ground, we will continue to come alongside with a research lens to present to the world what this strange new humanoid robot phenomenon is all about.

Acknowledgments

Conflicts of Interest

TB is the developer of the behaviors in MEDi software.

References

1. Claridge A, Hajec L, Montgomery L, Knapp B. Child and parent psychosocial experiences of hospitalization: an examination of the role of child life specialists. J of Child Life 2020;1(1):3-14.

2. Association of Child Life Professionals. 2020. URL: https://www.childlife.org/about-aclp [accessed 2020-06-26]

3. American Academy of Pediatrics Child Life Council Committee on Hospital Care, Wilson JM. Child life services. Pediatrics 2006 Oct;118(4):1757-1763. [doi: 10.1542/peds.2006-1941] [Medline: 17015572]

4. Dawe J, Sutherland C, Barco A, Broadbent E. Can social robots help children in healthcare contexts? A scoping review. BMJ Paediatr Open 2019;3(1):e000371 [FREE Full text] [doi: 10.1136/bmjpo-2018-000371] [Medline: 30815587]

5. Logan DE, Breazeal C, Goodwin MS, Jeong S, O'Connell B, Smith-Freedman D, et al. Social robots for hospitalized children. Pediatrics 2019 Jul;144(1):1-11. [doi: 10.1542/peds.2018-1511] [Medline: 31243158]

6. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. Vaccine 2013 Jun 7;31(25):2772-2777. [doi: 10.1016/j.vaccine.2013.03.056] [Medline: 23623861]

7. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Humanoid robotics in health care: an exploration of children's and parents' emotional reactions. J Health Psychol 2015 Jul;20(7):984-989. [doi: 10.1177/1359105313504794] [Medline: 24140615]

8. Beran T. Discovering How Humanoid Robotics Can Be Used for Pediatric Procedural Pain. In: 10th International Symposium on Pediatric Pain. 2015 Presented at: SPP'15; May 31-June 4, 2015; Seattle, USA.

9. Farrier C, Pearson JD, Beran TN. Children's fear and pain during medical procedures: a quality improvement study with a humanoid robot. Can J Nurs Res 2019 Jul 18:844562119862742. [doi: 10.1177/0844562119862742] [Medline: 31318580]

10. Jibb LA, Birnie KA, Nathan PC, Beran TN, Hum V, Victor JC, et al. Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: A pilot randomized controlled trial. Pediatr Blood Cancer 2018 Sep;65(9):e27242. [doi: 10.1002/pbc.27242] [Medline: 29893482]

11. Manaloor R. Humanoid robot-based distraction to reduce pain and distress during venipuncture in the pediatric emergency department: a randomized controlled trial. Paed Child Healt 2019;24(2):e43 [FREE Full text] [doi: 10.1093/pch/pxz066.112]

12. Meghdari A, Shariati A, Alemi M, Vossoughi GR, Eydi A, Ahmadi E, et al. Arash: A social robot buddy to support children with cancer in a hospital environment. Proc Inst Mech Eng H 2018 Jun;232(6):605-618. [doi: 10.1177/0954411918777520] [Medline: 29890934]

13. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Jan;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]

14. Robins B, Dautenhahn K, Dickerson P, Stribling P. Robot mediated joint attention in children with autism. Interact Stud 2004;5:198.

15. Pearson J. Integrating a humanoid robot into child life. Assoc Child Life Prof Bull 2017;35(2):13.

16. NAO - Technical overview. Softbank Robotics. URL: http://doc.aldebaran.com/2-1/family/robots/index_robots.html [accessed 2020-10-21]

17. Pearson J. The future is now: Using humanoid robots in child life practice. In: Handbook of Medical Play Therapy and Child Life: Interventions in Clinical and Medical Settings. New York, USA: Routedge; 2018:351-372.

18. Lal S, Suto M, Ungar M. Examining the potential of combining the methods of grounded theory and narrative inquiry: a comparative analysis. Qual Report 2012;17:1-22.

19. Ollerenshaw JA, Creswell JW. Narrative research: a comparison of two restorying data analysis approaches. Qual Inq 2016 Jun 29;8(3):329-347. [doi: 10.1177/10778004008003008]

20. Strauss A, Corbin J. Basics of Qualitative Research. London, UK: Sage Publications; 1990.

21. Thomas S, Pollio H. Listening to Patients: A Phenomenological Approach to Nursing Research and Practice. New York, USA: Springer; 2002.

22. Galletta A. Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication. New York, USA: New York University Press; 2013.

23. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

24. Corbin J, Strauss A. Basics of Qualitative Research. Third Edition. Thousand Oaks, CA: Sage; 2008.

25. Truog RD. Patients and doctors--evolution of a relationship. N Engl J Med 2012 Feb 16;366(7):581-585. [doi: 10.1056/NEJMp1110848] [Medline: 22335734]

26. Frampton S, Guastello S, Hoy L, Naylor M, Sheridan S, Johnston-Fleece M. Harnessing Evidence and Experience to Change Culture: a Guiding Framework for Patient and Family Engaged Care. National Academy of Medicine. 2017. URL: https://pdfs.semanticscholar.org/dbe1/e37ae998f1a67d99a3e49533cbf6d785d9b4.pdf [accessed 2020-07-07]

27. Looije R, Neerincx MA, Peters JK, Henkemans OA. Integrating robot support functions into varied activities at returning hospital visits. Int J of Soc Robotics 2016 Jun 23;8(4):483-497. [doi: 10.1007/s12369-016-0365-8]

28. Alemi M, Ghanbarzadeh A, Meghdari A, Moghadam LJ. Clinical application of a humanoid robot in pediatric cancer interventions. Int J of Soc Robotics 2015 Mar 18;8(5):743-759. [doi: 10.1007/s12369-015-0294-y]

29. Gjurković T, Tudor K. Treatment stages in working with children: an approach rooted in transactional analysis and play therapy. Transactional Analy J 2018 May 18;48(3):242-257. [doi: 10.1080/03621537.2018.1471291]

30. Schramm L, Dufault D, Young J. Warning: This Robot is Not What It Seems!. Cambridge, UK: 20 Companion; 2020.

Abbreviations

ASD: autism spectrum disorder
CLS: child life specialist
RA: research assistant

Original Paper

Perspectives of Patients, Health Care Professionals, and Developers Toward Blockchain-Based Health Information Exchange: Qualitative Study

Keehyuck Lee[1,2*], MD, MBA; Kahyun Lim[1*], MS; Se Young Jung[1,2], MD, MPH; Hyerim Ji[1], MS; Kyungpyo Hong[1], MS; Hee Hwang[1,2], MD, PhD; Ho-Young Lee[1,3], MD, PhD

[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[2]Department of Family Medicine, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[3]Department of Nuclear Medicine, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Se Young Jung, MD, MPH
Office of eHealth Research and Business
Seoul National University Bundang Hospital
Gyeonggi-do, Bundang-gu
Gumi-ro 173beon-gil
Seongnam-si, 13620
Republic of Korea
Phone: 82 317870114
Email: imsyjung@gmail.com

Abstract

Background: Although the electronic health record system adoption rate has reached 96% in the United States, implementation and usage of health information exchange (HIE) is still lagging behind. Blockchain has come into the spotlight as a technology to solve this problem. However, there have been no studies assessing the perspectives of different stakeholders regarding blockchain-based patient-centered HIE.

Objective: The objective of this study was to analyze the awareness among patients, health care professionals, and information technology developers toward blockchain-based HIE, and compare their different perspectives related to the platform using a qualitative research methodology.

Methods: In this qualitative study, we applied grounded theory and the Promoting Action on Research Implementation in the Health Service (PARiHS) framework. We interviewed 7 patients, 7 physicians, and 7 developers, for a total of 21 interviewees.

Results: Regarding the leakage of health information, the patient group did not have concerns in contrast to the physician and developer groups. Physicians were particularly concerned about the fact that errors in the data cannot be easily fixed due to the nature of blockchain technology. Patients were not against the idea of providing information for clinical trials or research institutions. They wished to be provided with the results of clinical research rather than being compensated for providing data. The developers emphasized that blockchain must be technically mature before it can be applied to the health care scene, and standards of medical information to be exchanged must first be established.

Conclusions: The three groups' perceptions of blockchain were generally positive about the idea of patients having the control of sharing their own health information. However, they were skeptical about the cooperation among various institutions and implementation for data standardization in the establishment process, in addition to how the service will be employed in practice. Taking these factors into consideration during planning, development, and operation of a platform will contribute to establishing practical treatment plans and tracking in a more convenient manner for both patients and physicians. Furthermore, it will help expand the related research and health management industry based on blockchain.

Introduction

Background

The Health Information Technology for Economic and Clinical Health (HITECH) Act was enacted in 2009 in the United States following evidence that the use of electronic health records (EHRs) can improve the quality of health care [1]. Since the law came into effect, the health care industry has seen a number of changes [2]. The EHR adoption rate has increased by 8% annually, which reached 96% of general medicine and surgical hospitals, 89% of rehabilitation hospitals, and 87% of children's hospitals in 2017 [3,4]. An EHR adoption program can achieve full success through health information exchange (HIE) because, similar to the exchange of information over the internet, medical information has the greatest value when exchange occurs without barriers [5]. However, according to statistics from the US government, significant progress must be made to reach a fully networked, patient-centered, integrated HIE system [6]. The situation in Korea is not much different. The government of the Republic of Korea has been developing an HIE system since 2006, but it is not yet fully operational because the information shared is only limited to medical institutions and not all data related to health care consumers' needs are provided [7-10].

Recently, blockchain has come into the spotlight as a technology to solve this problem [11-15]. Blockchain technology is a distributed ledger system that has no central authority, and the metadata are stored in "blocks," also known as on-chain data, which are securely created and are immutable owing to the use of unique cryptography [16]. Metadata on the blocks are linked to off-chain data in local hospital servers. Unlike traditional methods of storing and recording data from central servers, with blockchain technology the users validate, record, and manage the same data by themselves via person-to-person networks to ensure the trustworthiness, transparency, and security of the data. Owing to these benefits, blockchain technology is being adopted in various industries to foster innovation, and the health care industry is no exception.

A previous review paper explored the current status of health care worldwide, and indicated that proposed solutions impacted by blockchain are mostly EHRs, personal health records (PHRs), and clinical trial support systems [17]. The processes within the target systems were mainly focused on storage, sharing, and the exchange and access of health care data. However, uptake of the health care blockchain has been slow around the world [18]. The main reason is users' lack of motivation to use the blockchain platform actively. The lack of motivation in health care blockchain is mostly due to not recognizing its advantages such as the improvement to data security, privacy, integrity, interoperability, and authentication [19]. In terms of HIE, the advantage of this technology is that the individual has the authority to control the scope of the institution and information, while ensuring security. The disadvantage of blockchain is that the data are verified by many parties and the capacity of the block is limited, resulting in a slow and difficult task to process large volumes [20,21]. In addition, blockchain is known to be inflexible because once the information is recorded on the block,

it is very hard to delete and there is almost no way to determine whether the private key for personal authentication and decryption has been hacked or lost [22].

However, in a health care blockchain, patients' private keys can be stored in hospitals with well-established security systems and there are a few ways in which the above-mentioned disadvantages can be overcome. First, with only limited trusted parties such as hospitals and government institutions involved in the blockchain consensus algorithm by using a private blockchain, the verification process can be quicker [19]. Second, data that are too sensitive or large can be managed externally via the off-chain method, which resolves the inflexibility and difficulty in deleting data from the blockchain while increasing the efficiency in processing data of smaller size. In addition, in terms of HIE, there is a drawback due to the open nature of the blockchain, as patient information can be exposed to other parties not authorized by the patients [23]. This can be solved through the use of smart contracts in blockchain with dynamic consent, which allows the individual with the authority to control the parties that can access the information and to what extent, while ensuring security and data integrity [24].

Previous Research

A previous study revealed that HIE can reduce unnecessary medical expenses [8]. Despite the advantage of HIE, the adoption rates have still been low [25,26]. One study analyzed and reported the perceptions of medical staff on the health care blockchain [27]. Another study analyzed how to integrate data from mobile health care apps into the blockchain [28]. In addition, research has focused on how to manage PHR data in a blockchain [29], and how blockchain technology can be used to recruit patients and manage the entire process for clinical trials [30,31]. However, no studies have proven the effectiveness of applying blockchain technology directly to HIE, and in particular, there has been no research on how consumers, patients, or developers appreciate such blockchain-based medical information exchange based on a qualitative research method. Therefore, the aim of this study was to establish the grounds for developing a blockchain platform that can be utilized in the actual health care environment by analyzing the awareness and concerns of the stakeholders about blockchain health care based on qualitative research methodology.

Aim

Regarding the limitations in previous studies, the objective of this study was to analyze the awareness among patients, health care professionals, and information technology (IT) developers toward blockchain-based HIE, and to compare their different perspectives on the platform using qualitative research methodology. The results of this study can serve as the grounds for developing a blockchain platform that can be utilized in the actual health care environment.

Methods

Participants

This study was designed using a qualitative approach to analyze the awareness among various stakeholders toward sharing personal health information based on blockchain technology

that is patient-centered. The participants included patients (medical service consumers), physicians (health care providers), and technical developers. Semistructured interviews were conducted to collect and analyze their awareness and views of health care blockchain. No other specific inclusion criteria were set when selecting the participants. However, only the developers who had experience in developing health management systems, hospital information systems, or information exchange systems were selected to collect data to meet the purpose of the study.

Design

We applied grounded theory methodology, which is used when a conceptual framework of a phenomenon has not been defined clearly, understanding of the relationship between concepts is lacking, or when relevant and irrelevant variables cannot be determined because studies have not been conducted repeatedly on a certain topic [32]. Grounded theory uses inductive reasoning alongside qualitative data collection. In this study, we recorded the opinions of three types of stakeholders on blockchain-based HIE through an iterative process of discussion based on grounded theory. In addition, the interview questions were determined based on the Promoting Action on Research Implementation in the Health Service (PARiHS) framework [33], which was developed to provide insight on how the topics discussed in a study are implemented in the actual environment (ie, how to apply the knowledge into practice). This is largely based on three factors: literature or knowledge-based evidence (Evidence), consideration of situations (Context), and facilitating elements to introduce research details into practice (Facilitation) [34]. Since the broader aim of this study was to preliminarily develop a PHR system that incorporates blockchain-based HIE, we adopted the PARiHS framework to be able to transfer the results to the development of a PHR.

Based on this theory and framework, semistructured interview questionnaires were developed for each interview group. In this study, Evidence was subdivided into Awareness and Prior Experience; Context was subdivided into Existing Problems and Attitudes; and Facilitation was subdivided into Perceived Risk, Perceived Benefits, and Suggestions. Several questions were assigned to each subfactor. The list of detailed questions is shown in Multimedia Appendix 1. The interviewers explained the basic concept of blockchain during the interview process so that the participants could respond to the research questions with proper background knowledge. This study also followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [35].

Setting

The interviewed patients and physicians were recruited from Seoul National University Bundang Hospital, which has supervised a government-led pilot project on building an HIE system; thus, the health care professionals' understanding of IT and information exchange based on an IT system is fairly high at this hospital [10,36-39]. Furthermore, various IT infrastructures are provided at this hospital, such as guidance or a payment kiosk for patients, smart bedside stations, patient portals, and patient guide systems [40-48]. Therefore, the patients are familiar with the IT infrastructures at the hospital,

and the researchers decided that it would be easy to help the patients and health care professionals at this hospital clearly understand the concept of consumer-centered HIE based on blockchain technology.

For participant recruitment, purposive sampling and snowball sampling were adopted together. Purposive sampling was used for recruiting patients and physicians, whereas snowball sampling was used for recruitment of IT developers. Purposive sampling is a nonprobability sampling method in which researchers select a sample in accordance with their own judgment [49]. Snowball sampling is also a nonprobability sampling method in which the samples are recruited from the pool of participants by continuously being introduced to more potential study participants. This method is useful when it is difficult to find study participants who belong to a certain population [50]. Based on a previous study, interviews were designed to include at least 5 people per group [35].

We included 7 patients, physicians, and developers, respectively, for a total of 21 interviewees, and data saturation was confirmed while conducting the interviews. When selecting the participants, patients with chronic illness who visited the hospital often were selected primarily because they were already familiar with hospital processes. Furthermore, developers who had experience in developing PHR systems, hospital information systems, or information exchange systems were selected to collect data that met the purpose of the study. Physicians who were professors working at Seoul National University Bundang Hospital for at least 1 year and had experience in HIE were selected as interviewees.

Data Collection

Two researchers conducted face-to-face, semistructured interviews. One researcher who was working on a pilot project in developing nationwide HIE in South Korea led the interviews in the physician and developer groups. She is also a trained data scientist and health information analyst. The positive relationship with the participants in the two groups created an open atmosphere by encouraging participants to speak more freely about their experience. The other researcher led the interviews in the patients group. The interviews were recorded with the participants' consent and lasted about 20 to 40 minutes. There were no other individuals present except for the participants and interviewers. All participants signed the consent form voluntarily and were informed that they could freely withdraw from the study at any time. During the sessions, they referenced the semistructured interview questionnaire that covered all topics relevant to their experiences and opinions regarding a blockchain-based PHR. All sessions were conducted in a one-on-one manner.

The researchers followed the interview guidelines developed based on previous studies and were approved by members of the eHealth research team at Seoul National University Bundang Hospital.

Data Analysis

Explorative content analysis was performed to capture the opinions and perspectives on blockchain-based HIE through a PHR. The recorded interviews were transcribed by two

researchers who had been researching blockchain technology for more than 1 year, who crosschecked each other's work to produce the final transcript for analysis. Subsequently, each of the two researchers reviewed the transcription and then coded each subfactor. The coding manual was defined while inspecting the domain system together during categorization. The coding schemes developed by each researcher independently were then compared and analyzed according to the final coding manual, and the results were drawn after reviewing and agreeing upon the interview details.

In addition, the transcripts were further reviewed and corrected repeatedly by all coresearchers to enhance the accuracy of transcriptions. All researchers verified the results until reaching a consensus on clarified classifications.

Ethics

This study was approved by the Institutional Review Board of Human Research of Seoul National University Bundang Hospital, Republic of Korea (Protocol No. B-1910-571-302).

Results

Participants

The average age of the interviewees was 51 years (SD 15.7 years, range 15-72 years). Considering that the majority of patients at the hospital are middle-aged and elderly, middle-aged people were preferentially recruited since they are more IT-savvy. The demographics of the study participants are presented in Table 1.

Table 1. Demographic characteristics of participants (N=21).

Characteristic	n (%)
Interviewee Group	
Patients	7 (33)
Doctors	7 (33)
Developers	7 (33)
Gender	
Male	16 (76)
Female	5 (24)
Age (years)	
<30	0 (0)
30-39	4 (19)
40-49	11 (52)
50-59	5 (24)
60-69	0 (0)
≥70	1 (5)
Length of career (years)[a]	
1-4	1 (7)
5-10	4 (29)
11-15	5 (36)
16-20	3 (21)
≥21	1 (7)

[a]Doctors and developers only (N=14).

Evidence: Awareness of Blockchain, HIE, and PHR

The detailed responses to the questions for all three stakeholder groups are presented in Multimedia Appendix 2, and an overview of the responses is provided in Table 2. Most patients who were interviewed (5/7) did not know about PHR and HIE. In addition, none of the patients was aware of blockchain-based HIE. By contrast, all physicians and developers were aware of, had knowledge about the usage, or participated in the development of PHR, HIE, and health care blockchain.

Table 2. Summary of the interviews for each component of the Promoting Action on Research Implementation in the Health Service (PARiHS) framework.

PARiHS Component	Patients	Physicians	IT[a] developers
Evidence			
Awareness of health blockchain and the concepts of HIE[b] and PHR[c]	No	Yes	Yes
Prior experience with HIE	Yes, sharing information (self, parents, children)	Yes, sharing patient information	Yes, developed a related service
Context			
Existing problems in exchanging health information	Was not guided by the hospital	Too much unnecessary information; not all patients/hospitals participate	No problem in the system; the utilization is low
Attitudes about blockchain-based patient-centered HIE	Positive	Somewhat positive	Somewhat positive
Facilitation			
Perceived Risk	Difficult to use	Concerned about security issue; possibility of legal conflicts if all information is shared. Difficult to edit data. Health care professionals would become more conservative during treatment	Possibility of data loss due to users' inexperience. Data standardization needs to be done at each institution/corporation
Perceived Risk (information safety)	No	Yes	Yes
Suggestion			
Function	Guardian controlling information on behalf of a patient (eg, elderly parent, underage children). Information sharing network for rare blood type/diseases	Recruiting patients for clinical trials, receiving only the requested information	Distributed storage of actual data, setting managing entity for metadata
Utilization measures	Ease of use, education for usage, and recommendation by health care professionals	Government leadership at initial stabilization stage (incentive system, etc)	None
Data exchanging institutions	Hospital: Yes; Clinical Research Institute: Yes; Corporation: No; Health Management Center: indeterminate	Hospital: Yes; Clinical Research Institute: Yes; Corporation: indeterminate; Health Management Center: indeterminate	Hospital: Yes; Clinical Research Institute: Yes; Corporation: Yes; Health Management Center: Yes

[a]IT: information technology.

[b]HIE: hospital information exchange.

[c]PHR: personal health record.

All participants in the patients group had experience in sharing personal health information by submitting their health information for hospital, insurance claim, and health management purposes. Several participants also had experience in sharing health information of their parents or children. All participants reported that they physically visited the hospital to receive necessary documents. All patients mentioned that this was a complicated, time-consuming, and burdensome process. Nevertheless, the patients followed the process because it is a health-related issue, required for an insurance claim (ie, economical), and to avoid repeated medical checkups. Most of all, the patients reported that the main reason for selecting this route was because they were "told by the hospital to do so."

Physicians responded that they understood HIE and PHR in theory but did not frequently use such a system. Some IT developers responded that they did not have difficulty in developing the HIE or PHR services but faced difficulty in data mapping during the planning stage as they lacked understanding of the business (Table 3).

Table 3. Prior experience related to the awareness of blockchain, hospital information exchange (HIE), and personal health record (PHR) systems.

Group	Quotes
Patients	"My parents who live in Gangneung were ill, therefore they needed to be taken care of by the family. So, my parents and I transferred the records from a hospital in Gangneung to Bundang [where the participant lives]. I submitted the health records when making an insurance claim. I had to go to the hospital in Gangneung where my parents used to attend just to get the documents I needed. I submitted the documents for the insurance company through fax. It was too far, so I had to take a day off from work. It was too complicated and inconvenient. I had to see the doctor again to get issued documents, and could not get all the necessary documents from one place. I would have not done this if it was not for my parents' health."
	"After visiting a local gynecologic clinic, I had to go to a bigger hospital where I submitted examination results and referral notes. I had to submit health records when making an insurance claim. It's fine when I'm not busy, but it is a burden to get all these different kinds of documents when I'm busy. It would be better if they can give me all the documents at once, otherwise it's very inconvenient."
Physicians	"I'm aware of such a service, and I sometimes look up patient information on a health information exchange system. But I don't use it often."
	"Not all hospitals participate, and it didn't seem like the information I need is organized in a way so that I can easily find it. I don't use it often because I have many patients to checkup, so I don't want to waste time."
	"I find it quite useful. If the patient is referred from another hospital, it was good for checking their diagnosis and examination result."
Developers	"I was involved in the development process. Aside from the development itself, data mapping can proceed only if we understand the business. So, the hardest part was to discuss this with busy doctors and to make them understand it fully."

Context: Existing Problems in Exchanging Health Information

All patients had a positive response to PHR and HIE after it was explained to them. They responded that they had not been able to use such a system because they were not informed of a more convenient way. Although all participants use smartphones, the overall IT literacy is not high. Hence, they reported that looking up information on the internet and learning new methods by themselves is difficult. One patient who was aware of the PHR app and HIE service responded that the PHR app is merely used for checking information and the HIE service provided the information-sharing service between hospitals, and is not very useful since he does not visit any other hospital often. For physicians, the HIE system does not contain information that is actually needed, and an excessive amount of information is randomly provided and is thus not being used often during treatment. Further, patient information could not be found in some cases because not all hospitals were participating in the system. The developers responded that the system is well established although its utilization is low (Table 4).

Table 4. Opinions about existing problems in hospital information exchange (HIE).

Group	Quotes
Patients	"I was told to do so. Administrative procedures at the hospital seem pretty complicated, so I just did what they told me to do without any doubts. I didn't know about such a great system [HIE service] because no one told me. I would have used it if I knew…"
	"I really needed a service like this. I visit the hospital often, and so do the elderly in my family… But, do I need to go to the hospital to use it?"
	"I've heard of the health information exchange service, but I don't feel the need to use it because I don't switch hospitals often unless the information is shared somewhere else."
Physicians	"There is too much unnecessary information. The time is limited, and I get confused because some of the information is not what I need for treatment. It would be great if the exam results are related to the actual referral be highlighted."
	"In ophthalmology, blood test results in the system are not very useful for us. The referral written on paper actually has more relevant information."
	"Because not all hospitals participate in the system, sometimes patient information cannot be searched. Also, it's difficult to find out the situation of the institution I'm referring the patient to. For example, I'd like to refer a patient to get an abdominal ultrasound after 6 months, but I can't tell if that institution can perform abdominal ultrasound. When I'm referring a patient to several departments, I'm not sure if I need to fill out the information for each department separately or do it as a whole."
Developers	"No improvement is necessary in terms of the system. The biggest problem is that its utilization is low. Legal or systematic actions should be taken to standardize data exchange and make it mandatory."
	"It's not being used widely after development. Various cases need to be discovered to make advancements for higher effectiveness. I'm disappointed that the usage rate is so low."

Context: Attitudes on Blockchain-Based Patient-Centered HIE

Patients were positive about the idea of sharing personal information not only with hospitals but also with insurance companies or health management service providers under their control. Many patients receive diverse treatments and manage their health in addition to visiting hospitals. For example, some patients with lower back pain visited orthopedics and traditional medicine clinics, and engaged in personal training exercises at the same time, in which they provided X-ray results or diagnostic information to traditional medicine doctors and personal trainers. They were particularly positive about how they could provide any information to these institutions on their own will using smartphones.

Physicians were also very positive about the idea of requesting and receiving only the information they or their patients need, unlike the existing HIE service in which data formatted in a set Clinical Document Architecture format must be sent or received. Developers identified the purpose of such a system or service but pointed out that the data managed independently by various institutions and companies need to be standardized to enable searching of and sharing certain data using blockchain, ultimately implementing interoperability in performing actual tasks. However, they responded that the most challenging part would be having institutions and companies voluntarily follow these processes and participate in the system. Once this issue is resolved, they explained that there will be no technical difficulty in creating blockchain-based user authentication or information exchange systems (Table 5).

Table 5. Attitudes on blockchain-based patient-centered hospital information exchange (HIE).

Group	Quotes
Patients	"So now I can just use my phone to send my information to wherever I want without having to visit the hospital after going through the consent process. It's less work for everyone. Not just for hospitals but it's easier to send documents to insurance companies as well."
	"I think it's great. If I happen to end up at an emergency room, the doctors would know what kind of illness or allergies I have?"
	"Because of my lower back pain, I visit an orthopedic doctor but also a traditional medicine clinic near where I live. The doctor at the traditional medicine clinic asked for my X-ray when I visited. Now I can show him my X-ray without having to visit the hospital I usually go to?"
Physicians	"I think that the most ideal way is to exchange patient-centered medical information through blockchain."
	"I would like to participate in the blockchain-based medical information exchange platform, but I suggest that the information be standardized before that."
Developers	"Blockchain is expected to be used a lot in the health care fields. Right now, the PHR [personal health record] only has a simple prescription record and a diagnostic record, but I think it would be helpful for you at the enterprise level if you could obtain a lot of data related to health care."

Facilitation: Perceived Risk

Patients were very positive about the blockchain-based patient-centered HIE; however, they were reluctant or even afraid to search for new information and learn new methods as blockchain is a relatively new technology. Among the 7 patients who were interviewed, 6 patients responded that they did not try to learn new IT. However, they were willing to learn and use the new system if their primary care physician or the hospital provided instructions and education. They were not particularly concerned with a leakage of health information, unlike physicians, who were sensitive to information security issues.

There are cases of errors made in patient charts, and physicians were concerned with being unable to revise or edit information in blockchain. If all information can be accessed by anyone upon patient consent, the information could be shared with lawyers at any time, which would create an environment in which legal conflicts regarding medical practices may occur more frequently. The physicians explained that the system would cause them to be more conservative during treatment. Some also emphasized that it would encourage doctor shopping even more. Developers also considered security as the greatest risk. Specifically, the developers thought that if information gathered on the blockchain platform was hacked or released accidentally, all medical information could be exposed. They also expressed concern about the situation in which users could not access their medical information at all if they inadvertently lost their private keys. The developers responded that blockchain has been proven only theoretically so far, and its reliability of information security in the real world has not been verified sufficiently through experiences (Table 6).

Table 6. Perceived risks of a blockchain-based hospital information exchange (HIE) system.

Group	Quotes
Patients	"I'm not too concerned about the security. I don't think it would be a huge risk for me even if my health information is leaked. An old saying goes, 'the more serious the illness, the more you should tell people'."
	"Because of my age, I'm not too interested in looking for new ways to do things and learn new things. But I'm not incapable of doing all that. If it's something that must be done, I'd go through it, even if it's a little burdensome."
Physicians	"Considering the situation in Korea, doctor shopping [the practice of visiting different hospitals to get a diagnosis because they don't trust a doctor's diagnosis for their pathological symptoms] would happen more commonly because of the convenient HIE system."
	"Being unable to revise the data would be a problem. We are human, and everyone makes mistakes when generating data. But there's no opportunity to fix errors. If the transaction is made again to fix the error, patients would know that their records contain errors, which would cause conflicts between the patient and the hospital. In ophthalmology, the records of left and right eyes are switched from time to time."
	"If anybody can see the treatment history, even small details would be available for lawyers to see and there's a possibility of being dragged into conflict for everything. The ideal case would be when data are used for positive purposes such as research or health management, but health care professionals would actually become very defensive when diagnosing patients and making care plans."
	"Some patients might refuse to provide data that are absolutely necessary, and it might become even harder when the physicians exchange patient information."
Developers	"If consent is given through apps, there is a risk of health data being leaked when a mobile phone is lost or an app is hacked. In cryptocurrency systems, people cannot withdraw their money if their personal key or mobile phone is lost. The fact that the patient cannot find their own health information should be taken very seriously."
	"Blockchain technology has not been around for long and is still being researched, there are still questions about the reliability of blockchain technology. In order to gain reliability of this technology, small-scale projects for building clinical study platforms should be conducted more and carried out up to actual demonstration."
	"If data are lost due to carelessness or inexperience, restoring the lost data would be difficult. Health data can be disclosed if a mobile phone is lost or an app is hacked."

Suggestion: Function

Most patients wished they could control the information of their parents or children rather than their own information. They also suggested that health care professionals could have access to their information without their consent in an emergency. They further suggested saving rare blood type, allergies, and organ donor information on the system and sharing such information when needed. Physicians anticipated that recruiting patients for clinical trials would become easier. All 7 physicians who were interviewed had difficulties with recruiting patients for clinical trials. Even when patients are recruited through advertisement, statistical distribution cannot be taken into consideration. They regarded blockchain as a solution for solving this problem to some extent. At the same time, however, they were not comfortable with having access to too much information. They responded that the technology needs to consider how it will show only the information that physicians actually need. Developers were concerned with data storage. An extensive amount of data cannot be stored on blockchain, which may cause a significant drop in the speed. All of the developers agreed that it is appropriate for actual data to be stored at each institution, whereas data generation, search, and exchange should be available through blockchain (Table 7).

XSL•FO

RenderX

Table 7. Suggested functions of a blockchain-based hospital information exchange (HIE) system.

Group	Quotes
Patients	"Those who really need this service are my parents and they have dementia. It may not always be the right decision to give authorization for sharing information to oneself. Sometimes it's better to ask for consent from the guardian."
	"My elderly patients are scared of visiting hospitals alone. It'll be great if I have the authorization to control my parents' information and I won't have to visit the hospital as often."
	"Information on rare blood type, allergies, and organ donors should be stored and shared whenever needed."
	"I wanted to participate in clinical trials, but no information was available. I wish someone would tell me."
	"Technically I get to control my own information, but doctors should be able to look up my information in an emergency. If I get into an accident overseas and become unconscious, it'll be helpful if my guardian or doctors could see my information."
Physicians	"Recruiting clinical trial participants is very difficult, and I think that will be most useful when using blockchain in health care. Recruiting participants appropriate for each stage and conducting follow up are the hardest parts."
	"Currently, recruiting patients for clinical trials is done by rule of thumb. Gender or age distribution is not considered most of the time. It'll be much more convenient and effective if we can recruit participants by looking up patient information using blockchain."
	"Because we get to see patients very briefly, there has to be a way of clearly showing only the information we need. For example, highlighting only the abnormal values or generating graphs of data for which the changes need observation."
Developers	"Patient data will be generated from so many places. Who will be responsible for managing such a vast amount of data? It'll be most convenient if the data are stored in the central location at the national level, but it's not possible for the central government to manage PGHD [patient-generated health data]. It's more efficient to store metadata in the central location, and the data itself should be saved on users' mobile phone or other storage media."
	"I think the data should be distributed for storage, and a system that manages the history data in the central location to enable searching is needed."

Institutions to Which Participants Are Willing to Provide Information

The interviewees were asked for what purpose and to which institutions they would be willing to provide their information. All patients were willing to provide their information to hospitals for research purposes. They did not particularly want to be compensated for providing data, but they wished to get the results if they participate in clinical trials. Among the 7 patient interviewees, only 2 were willing to provide data to profit-based corporations. Some physicians were willing to provide data to all institutions, only for public interest, or to places that give greater compensation. Most of the developers responded that they would provide information to all institutions for appropriate compensation.

Discussion

Principal Findings

This study was conducted in an IT-friendly tertiary hospital equipped with a government-led nationwide HIE system, and we analyzed the various expectations and concerns that doctors, patients, and developers already exposed to HIE and a PHR environment had about a potential patient-oriented blockchain-based HIE platform. This study is the first to analyze three different perspectives by interviewing all three stakeholder groups that will actually develop and use a patient-centered blockchain-based HIE.

Regarding the leakage of health information, the patient group did not have concerns, whereas the physician and developer groups expressed such concerns. We actually expected patients to be sensitive with regard to the security of their health information; however, in contrast to these expectations, they were not particularly concerned. Normally, patients attending tertiary hospitals tend to trust hospitals and doctors in South Korea. This may the reason for the lack of concern of privacy issues found in this study. Another reason is that there have been few accidents of health care information leakage so far in South Korea. By contrast, physicians and developers consider the security of health information as one of the greatest values because they directly manage a hospital information system and understand how vulnerable the system is to cyberattack. Patients are unaware of the risks resulting from the leakage of health information, whereas those who handle medical records often are more aware of the risks of leakage, thus creating a discrepancy between the patient group and physician group.

Moreover, physicians were especially concerned about the fact that errors in the data cannot be easily fixed due to the nature of blockchain technology. Instead, using blockchain technology, when medical records need to be revised, a letter of explanation must be completed and approved after which relevant departments that may be affected by the revision have to be notified (at most medical institutions). Accordingly, the revision history is uploaded to the blockchain network. However, with respect to legal conflicts arising due to the sharing of data, only the metadata of health data are stored on the blockchain network while the actual data are rarely carried by the blockchain due to speed and size issues, thus preventing all of the data from being accessed by legal personnel. The hospitals are currently obligated to provide any requested data if patients want to take legal actions against health care professionals, which indicates that the issue of rising legal conflicts will not become worse. Developers were especially concerned with the difficulty in bringing all institutions and corporations together to agree on data standardization if they were to participate in creating a network, and in actually proceeding with the standardization process. The most ideal case would be to have the central government manage all of the data; however, the study

participants agreed that sharing metadata on the blockchain network and storing the actual data in the storage media of each institution or corporation are the most suitable methods, since concentration or abuse of power may occur from having exclusive rights over the data if all data are managed by a single entity. Data standardization is necessary to facilitate data exchange between different institutions. Terminology mapping must be performed for data standardization; however, this may entail an extensive amount of time and costs if detailed understanding of the business is lacking.

Patients have been frequently exchanging their health information for a variety of purposes. They were very positive about having control of their own information, with advantages of not having to visit the hospital to receive documents that need to be submitted to different hospitals, insurance companies, or other health management institutions (eg, chiropractic office, fitness center), and could avoid repeated examinations. However, most patients were afraid of and reluctant to accepting and learning new technology by themselves. They also have high trust in physicians and the guidance provided by the hospital. Most patients responded that they would like to start using new services if their doctors explain the need for using the new service while the hospital provides guidance. Furthermore, they expected doctors to be able to look up their information without their consent in emergencies, and to have control over the health information of their elderly parents or underage children. They were not against the idea of providing information for the purpose of participating in clinical research, and they wished to be provided with the results instead of compensation for providing the data. In contrast to our expectation, the patients were not particularly concerned with the leakage of their health information; because most of them lived in a different city from their elderly parents, they prefer being able to share their parents' health information using a mobile phone, instead of having to visit the hospital every time with their parents simply to explain their conditions or to receive necessary documents for insurance claims.

Regarding this issue, all physicians agreed that the results of clinical research should be shared with the participants; however, some disagreed as to what extent the details should be shared. Some argued that it is the right of the participants to be provided with the complete results of clinical research, whereas others suggested maintaining the passive stance in which information is provided only when requested by the participant. Some physicians explained that everyone needs to agree on when and to what extent they should share research results in addition to the results of the examinations conducted when participating in clinical research. The results cannot be shared until the findings are officially approved and conclusions are made; hence, sharing results is meaningless, as argued by some physicians.

Regarding service utilization, the patients were willing to use the new service if recommended by physicians and supported by the guidance and instructions from the hospital. Physicians emphasized the importance of the government's role in establishing an initial system and the verification process. The government should play an active role in establishing a compensation and incentive policy, provide a subsidy when

building computer power—which is required for system maintenance—and issuing coins for forming an initial ecosystem.

Blockchain and Patient-Centered HIE

According to the results of this study, all three stakeholders agreed that HIE is needed. From the physicians' perspective, the blockchain-based patient-centered HIE is not a problem of technology. The main driver is the government's policy to implement the system. For instance, the Japanese government is already trying to implement and evaluate several new ITs, including blockchains, to actively explore innovative and disruptive technologies through the Cabinet Secretariat's regulatory sandbox system [51]. Further, from the developers' perspective, there is no problem with respect to the blockchain technology itself. Rather, the problem lies in what type of information can be managed by the system. Due to the Personal Information Protection Act such as the EU General Data Protection Regulation (GDPR), Organisation for Economic Co-operation and Development (OECD) Privacy Guidelines, and Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules in the United States, people can request their own data to be deleted. This is possible when the recording itself is not stored on the blockchain [52]. Therefore, it is necessary to determine the extent of the data recorded on the blockchain, such as the hash value of the recorded health data, which cannot be linked to personal information, or data that will not be a risk even if accessed. Moreover, the challenge is to justify the implementation of the blockchain in HIE. If there are incentives for data tampering and using a trusted third party is expensive, using blockchain is a reasonable solution [53,54]. Health care blockchain is not suitable for the current health care environment given the low HIE adoption rate. However, in the future, HIE will play a similar role as the internet does today, and thus HIE is expected to be used daily. In this environment, blockchain can be a powerful driver for ensuring information privacy and security.

Limitations and Future Research

There are several limitations of this study. First, 21 participants cannot represent the opinion of the entire population, and the three groups of stakeholders cannot represent all stakeholders. Second, the age of the patient group was between 40 and 60 years, which could introduce bias in the responses. Third, patients and physicians were recruited from a single medical institution as the interviewees, thus failing to secure diversity in samples. More studies should be conducted in the future regarding these limitations. Finally, physicians and patients did not have technical knowledge of blockchain. However, we tried to provide proper background knowledge about blockchain before each interview so that the results of the interview were not biased.

Conclusion

We analyzed the opinions of stakeholders regarding a blockchain-based, patient-centered HIE platform based on interviews with physicians, IT developers, and patients using the PARiHS framework. Most of the participants were positive about the idea of patients having the control of sharing their

XSL•FO

RenderX

own health information, but were skeptical about the cooperation among various institutions and the implementation for data standardization in the establishment process, in addition to actual measures for utilization of the service. Taking these factors into consideration during planning, development, and operation of a platform will contribute to establishing practical treatment plans and tracking in a more convenient manner for both patients and physicians, and will help to expand the related research and health management industry.

Acknowledgments

Authors' Contributions

KL drafted the entire manuscript. HJ and KL conducted the interviews. KH and KL analyzed the data and drafted the Methods and Results sections. HH and HL contributed to the discussion and interpretation of data. SJ supervised the entire process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of interview questions.
[DOCX File , 17 KB - jmir_v22i11e18582_app1.docx]

Multimedia Appendix 2
Detailed responses to interview questions in each stakeholder group.
[DOCX File , 22 KB - jmir_v22i11e18582_app2.docx]

References

1. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med 2009 Apr 09;360(15):1477-1479. [doi: 10.1056/NEJMp0901592] [Medline: 19321856]
2. Gold M, McLaughlin C. Assessing HITECH Implementation and Lessons: 5 Years Later. Milbank Q 2016 Sep 13;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
3. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
4. Percent of Specialty Hospitals that Possess Certified Health IT. The Office of the National Coordinator for Health Information Technology. URL: https://dashboard.healthit.gov/quickstats/pages/specialty-hospital-ehr-adoption.php [accessed 2020-02-13]
5. Wu H, Larue E. Barriers and Facilitators of Health Information Exchange (HIE) Adoption in the United States. In: IEEE Explore. 2015 Jan 5 Presented at: 48th Hawaii International Conference on System Sciences; 2015; Hawaii p. 2015-2048. [doi: 10.1109/hicss.2015.356]
6. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb 21;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]
7. Park H, Lee S, Kim Y, Heo E, Lee J, Park JH, et al. Patients' perceptions of a health information exchange: a pilot program in South Korea. Int J Med Inform 2013 Feb;82(2):98-107. [doi: 10.1016/j.ijmedinf.2012.05.001] [Medline: 22658777]
8. Park H, Lee S, Hwang H, Kim Y, Heo E, Kim J, et al. Can a health information exchange save healthcare costs? Evidence from a pilot program in South Korea. Int J Med Inform 2015 Sep;84(9):658-666. [doi: 10.1016/j.ijmedinf.2015.05.008] [Medline: 26048738]
9. Lee SI, Park H, Kim JW, Hwang H, Cho EY, Kim Y, et al. Physicians' perceptions and use of a health information exchange: a pilot program in South Korea. Telemed J E Health 2012 Oct;18(8):604-612 [FREE Full text] [doi: 10.1089/tmj.2011.0261] [Medline: 22352898]
10. Lee M, Heo E, Lim H, Lee JY, Weon S, Chae H, et al. Developing a common health information exchange platform to implement a nationwide health information network in South Korea. Healthc Inform Res 2015 Jan;21(1):21-29 [FREE Full text] [doi: 10.4258/hir.2015.21.1.21] [Medline: 25705554]
11. Mettler M. Blockchain technology in healthcare: The revolution starts here. 2016 Sep 14 Presented at: 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom). . 1?3; 2016; Munich, Germany p. 2016. [doi: 10.1109/healthcom.2016.7749510]

12. Ekblaw A, Azaria A, Halamka J, Lippman A. A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data. 2016 Dec 5 Presented at: 2016 IEEE International Conference on Big Data; December 5-8, 2016; Washington, DC.

13. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare Blockchain System Using Smart Contracts for Secure Automated Remote Patient Monitoring. J Med Syst 2018 Jun 06;42(7):130. [doi: 10.1007/s10916-018-0982-x] [Medline: 29876661]

14. Zhang P, Schmidt DC, White J, Lenz G. Chapter One - Blockchain technology use cases in healthcare. Adv Comp 2018 Jan 1;111:1-41. [doi: 10.1016/bs.adcom.2018.03.006]

15. Amoedo A, Martnez-Costa MDP, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. Radio J Int Stud Broadcast Audio Media 2009 Feb 01;6(1):5-20. [doi: 10.1386/rajo.6.1.5_4]

16. Pilkington M. Blockchain technology: principles and applications. In: Olleros FX, Zhegu M, editors. Research Handbook on Digital Transformations. Northampton, MA: Edward Elgar Publishing; Jan 1, 2016:225-253.

17. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]

18. Budman M, Hurley B, Khan A, Gangopadhhyay N. Deloitte's 2019 global blockchain survey. Deloitte Insights. URL: https://www2.deloitte.com/content/dam/Deloitte/se/Documents/risk/DI_2019-global-blockchain-survey.pdf [accessed 2021-11-04]

19. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB. A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application: Taxonomy, Substantially Analysis, Motivations, Challenges, Recommendations and Future Direction. J Med Syst 2019 Sep 14;43(10):320. [doi: 10.1007/s10916-019-1445-8] [Medline: 31522262]

20. Croma C, Decker C, Eyal I, Gencer AE, Juels A, Kosba A, et al. On Scaling Decentralized Blockchains (A Position Paper). In: Lecture Notes in Computer Science volume 9604. 2016 Aug 31 Presented at: International Conference on Financial Cryptography and Data Security; February 22-26, 2016; Christ Church, Barbados p. 106-125. [doi: 10.1007/978-3-662-53357-4_8]

21. El-Gazzar R, Stendal K. Blockchain in Health Care: Hope or Hype? J Med Internet Res 2020 Jul 10;22(7):e17199 [FREE Full text] [doi: 10.2196/17199] [Medline: 32673219]

22. Zheng W, Wang K, Wang FY. GAN-Based Key Secret-Sharing Scheme in Blockchain. IEEE Trans Cybern 2020 Jan 22;26(4):1-8. [doi: 10.1109/TCYB.2019.2963138] [Medline: 31985449]

23. Angraal S, Krumholz HM, Schulz WL. Blockchain Technology: Applications in Health Care. Circ Cardiovasc Qual Outcomes 2017 Sep;10(9):e003800. [doi: 10.1161/CIRCOUTCOMES.117.003800] [Medline: 28912202]

24. Dimitrov DV. Blockchain Applications for Healthcare Data Management. Healthc Inform Res 2019 Jan;25(1):51-56 [FREE Full text] [doi: 10.4258/hir.2019.25.1.51] [Medline: 30788182]

25. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]

26. Yaraghi N, Du AY, Sharman R, Gopal RD, Ramesh R. Health Information Exchange as a Multisided Platform: Adoption, Usage, and Practice Involvement in Service Co-Production. Inf Syst Res 2015 Mar;26(1):1-18. [doi: 10.1287/isre.2014.0547]

27. Hau YS, Lee JM, Park J, Chang MC. Attitudes Toward Blockchain Technology in Managing Medical Information: Survey Study. J Med Internet Res 2019 Dec 09;21(12):e15870 [FREE Full text] [doi: 10.2196/15870] [Medline: 31815676]

28. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. 2017 Oct 8 Presented at: IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). . 1?5; 2017; Montreal p. 2017. [doi: 10.1109/pimrc.2017.8292361]

29. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility. J Med Internet Res 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]

30. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917. [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]

31. Maslove DM, Klein J, Brohman K, Martin P. Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]

32. Corbin JM, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21. [doi: 10.1007/bf00988593]

33. Stetler CB, Damschroder LJ, Helfrich CD, Hagedorn HJ. A Guide for applying a revised version of the PARIHS framework for implementation. Implement Sci 2011 Aug 30;6:99 [FREE Full text] [doi: 10.1186/1748-5908-6-99] [Medline: 21878092]

34. Laycock A, Harvey G, Percival N, Cunningham F, Bailie J, Matthews V, et al. Application of the i-PARIHS framework for enhancing understanding of interactive dissemination to achieve wide-scale improvement in Indigenous primary healthcare. Health Res Policy Syst 2018 Nov 29;16(1):117 [FREE Full text] [doi: 10.1186/s12961-018-0392-z] [Medline: 30497480]

35. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

36. Lee K, Jung SY, Hwang H, Yoo S, Baek HY, Baek R, et al. A novel concept for integrating and delivering health information using a comprehensive digital dashboard: An analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage. Int J Med Inform 2017 Jan;97:98-108. [doi: 10.1016/j.ijmedinf.2016.10.001] [Medline: 27919400]

37. Cho M, Kim K, Lim J, Baek H, Kim S, Hwang H, et al. Developing data-driven clinical pathways using electronic health records: The cases of total laparoscopic hysterectomy and rotator cuff tears. Int J Med Inform 2020 Jan;133:104015. [doi: 10.1016/j.ijmedinf.2019.104015] [Medline: 31683142]

38. Yoo S, Lim K, Baek H, Jang S, Hwang G, Kim H, et al. Developing a mobile epilepsy management application integrated with an electronic health record for effective seizure management. Int J Med Inform 2020 Feb;134:104051. [doi: 10.1016/j.ijmedinf.2019.104051] [Medline: 31837500]

39. Yoo S, Kim S, Lee K, Jeong CW, Youn SW, Park KU, et al. Electronically implemented clinical indicators based on a data warehouse in a tertiary hospital: its clinical benefit and effectiveness. Int J Med Inform 2014 Jul;83(7):507-516. [doi: 10.1016/j.ijmedinf.2014.04.001] [Medline: 24867195]

40. Baek M, Koo BK, Kim BJ, Hong K, Kim J, Yoo S, et al. Development and Utilization of a Patient-Oriented Outpatient Guidance System. Healthc Inform Res 2016 Jul;22(3):172-177 [FREE Full text] [doi: 10.4258/hir.2016.22.3.172] [Medline: 27525158]

41. Yoo S, Jung SY, Kim S, Kim E, Lee K, Chung E, et al. A personalized mobile patient guide system for a patient-centered smart hospital: Lessons learned from a usability test and satisfaction survey in a tertiary university hospital. Int J Med Inform 2016 Jul;91:20-30. [doi: 10.1016/j.ijmedinf.2016.04.003] [Medline: 27185506]

42. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]

43. Yoo S, Lee K, Baek H, Ryu B, Chung E, Kim K, et al. Development and User Research of a Smart Bedside Station System toward Patient-Centered Healthcare System. J Med Syst 2015 Sep;39(9):86. [doi: 10.1007/s10916-015-0273-8] [Medline: 26208595]

44. Ro HJ, Jung SY, Lee K, Hwang H, Yoo S, Baek H, et al. Establishing a Personal Health Record System in an Academic Hospital: One Year's Experience. Korean J Fam Med 2015 May;36(3):121-127 [FREE Full text] [doi: 10.4082/kjfm.2015.36.3.121] [Medline: 26019761]

45. Jung SY, Lee K, Hwang H, Yoo S, Baek HY, Kim J. Support for Sustainable Use of Personal Health Records: Understanding the Needs of Users as a First Step Towards Patient-Driven Mobile Health. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e19 [FREE Full text] [doi: 10.2196/mhealth.6021] [Medline: 28232300]

46. Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an Electronic Health Record-Integrated Personal Health Record on Patient Participation in Health Care: Development and Randomized Controlled Trial of MyHealthKeeper. J Med Internet Res 2017 Dec 07;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]

47. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of Personal Health Records and Wearables on Health Outcomes and Patient Response: Three-Arm Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]

48. Jung SY, Kim J, Hwang H, Lee K, Baek R, Lee H, et al. Development of Comprehensive Personal Health Records Integrating Patient-Generated Health Data Directly From Samsung S-Health and Apple Health Apps: Retrospective Cross-Sectional Observational Study. JMIR Mhealth Uhealth 2019 May 28;7(5):e12691 [FREE Full text] [doi: 10.2196/12691] [Medline: 31140446]

49. Gentles SJ, Charles C, Ploeg J, McKibbon KA. Sampling in qualitative research: Insights from an overview of the methods literature. Qual Rep 2015;20(11):1772-1789.

50. Noy C. Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. Int J Soc Res Method 2008 Oct;11(4):327-344. [doi: 10.1080/13645570701401305]

51. Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the Potential of Blockchain Technology to Meet the Needs of 21st-Century Japanese Health Care: Viewpoint on Use Cases and Policy. J Med Internet Res 2020 Jan 09;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]

52. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]

53. Wüst K, Gervais A. Do you need a blockchain? 2018 Jun 20 Presented at: 2018 Crypto Valley Conference on Blockchain Technology (CVCBT); June 20, 2018; Zug, Switzerland p. 2018 URL: https://ieeexplore.ieee.org/abstract/document/8525392/ [doi: 10.1109/CVCBT.2018.00011]

54. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative research
EHR: electronic health record
HIE: health Information exchange
HITECH: Health Information Technology for Economic and Clinical Health
IT: information technology
PARiHS: Promoting Action on Research Implementation in the Health Service
PHR: personal health record

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study

Hongyi Wang[1,2,3,4], PhD; Jing Zhang[1,2,3,4], PhD; Zhenxing Chu[1,2,3,4]; Qinghai Hu[1,2,3,4], Prof Dr; Willa Dong[5], MSPH; Xiaojie Huang[6], Prof Dr; Yaokai Chen[7], Prof Dr; Hui Wang[8], Prof Dr; Xiaoqing He[7]; Lukun Zhang[8], PhD; Zhili Hu[1,2,3,4]; Rantong Bao[1,2,3,4]; Shangcao Li[1,2,3,4]; Hang Li[1,2,3,4]; Sitong Cui[1,2,3,4]; Xia Jin[1,2,3,4]; Haibo Ding[1,2,3,4], PhD; Wenqing Geng[1,2,3,4], Prof Dr; Yongjun Jiang[1,2,3,4], Prof Dr; Junjie Xu[1,2,3,4*], Prof Dr; Hong Shang[1,2,3,4*], Prof Dr

[1]NHC Key Laboratory of AIDS Immunology (China Medical University), National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning Province, China

[2]Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, China

[3]Key Laboratory of AIDS Immunology of Liaoning Province, Shenyang, China

[4]Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, China

[5]Department of Health Behavior, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Center for Infectious Diseases, Beijing Youan Hospital, Capital Medical University, Beijing, China

[7]Chongqing Public Health Medical Center, Chongqing, China

[8]Shenzhen Third People's Hospital, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Hong Shang, Prof Dr
NHC Key Laboratory of AIDS Immunology (China Medical University), National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University
No 155, Nanjing North Street, Heping District
Shenyang, Liaoning Province, 110001
China
Phone: 86 8328 2634
Email: hongshang100@hotmail.com

Related Article:

Correction of: http://www.jmir.org/2020/10/e22388/

In "Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study" (J Med Internet Res 2020;22(10):e22388) the authors noted one error.

A typographical error in the originally published version of the paper rendered the second half of the article title as:

Real-Word, Multicenter Study

This has been corrected to:

Real-World, Multicenter Study

The correction will appear in the online version of the paper on the JMIR Publications website on November 9, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

<u>Corrigenda and Addenda</u>

Correction: Elaborating Models of eHealth Governance: Qualitative Systematic Review

Anne Granstrom Ekeland[1*], PhD; Line Helen Linstad[1*], MSc

Norwegian Centre for E-health Research, Tromsø, Norway

[*]all authors contributed equally

Corresponding Author:
Anne Granstrom Ekeland, PhD
Norwegian Centre for E-health Research
P.o. box 35, University Hospital
Tromsø, 9038
Norway
Phone: 47 95266791
Email: anne.granstrom.ekeland@ehealthresearch.no

Related Article:

Correction of: https://www.jmir.org/2020/10/e17214/

In "Elaborating Models of eHealth Governance: Qualitative Systematic Review" (J Med Internet Res 2020;22(10):e17214), the authors noted one error.

In the Methods section of the Abstract, the following sentence specified an incorrect date range:

We searched the PubMed database using predefined search terms and selected papers published in 2010.

This sentence has been corrected to:

We searched the PubMed database using predefined search terms and selected papers published from 2010 onwards.

The correction will appear in the online version of the paper on the JMIR Publications website on November 24, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO

RenderX

Corrigenda and Addenda

Correction: Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis

Anjana Estelle Sharma[1,2], MAS, MD; Ziva Mann[3], MA; Roy Cherian[2,4], MHS; Jan Bing Del Rosario[1,5], BS; Janine Yang[1,6], BS; Urmimala Sarkar[2], MD, MPH

[1]Department of Family & Community Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Center for Vulnerable Populations, University of California San Francisco, San Francisco, CA, United States

[3]Ziva Mann Consulting, Newton, MA, United States

[4]Department of Culture and Theory, School of Humanities, University of California, Irvine, Irvine, CA, United States

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

[6]Drexel University College of Medicine, Philadelphia, PA, United States

Corresponding Author:
Anjana Estelle Sharma, MAS, MD
Department of Family & Community Medicine
University of California San Francisco
995 Potrero Ave
Ward 83
San Francisco, CA, 94110
United States
Phone: 1 6179454776
Email: anjana.sharma@ucsf.edu

Related Article:

Correction of: https://www.jmir.org/2020/10/e17595/

In "Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis (J Med Internet Res 2020;22(10):e17595), the authors noted three errors.

In the originally published manuscript, the fourth sentence of the second paragraph of the Introduction section read:

> *The term originated from a professional YouTube video maker, who posted a video on Twitter explaining that she had been diagnosed with Ehlers-Danlos syndrome and postural orthostatic hypotension syndrome (POTS).*

This has been changed to:

> *The term originated from a professional YouTube video maker, who posted a video on Twitter explaining that she had been diagnosed with Ehlers-Danlos syndrome and postural orthostatic tachycardia syndrome (POTS).*

In Table 2, footnote [a] originally read:

> *POTS: postural orthostatic hypotension syndrome.*

This has been changed to:

> *POTS: postural orthostatic tachycardia syndrome.*

The Abbreviations section originally included the following:

> **POTS:** *postural orthostatic hypotension syndrome.*

This has been changed to:

> **POTS:** *postural orthostatic tachycardia syndrome.*

These corrections will appear in the online version of the paper on the JMIR Publications website on November 9, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Original Paper

openEHR Archetype Use and Reuse Within Multilingual Clinical Data Sets: Case Study

Heather Leslie[1,2], MBBS, FAIDH, FIAHSI

[1]Atomica Informatics, Fitzroy, Australia

[2]openEHR International, Clinical Models Program, London, United Kingdom

Corresponding Author:
Heather Leslie, MBBS, FAIDH, FIAHSI
Atomica Informatics
Fitzroy
Australia
Phone: 61 418966670
Email: heather.leslie@atomicainformatics.com

Abstract

Background: Despite electronic health records being in existence for over 50 years, our ability to exchange health data remains frustratingly limited. Commonly used clinical content standards, and the information models that underpin them, are primarily related to health data exchange, and so are usually document- or message-focused. In contrast, over the past 12 years, the Clinical Models program at openEHR International has gradually established a governed, coordinated, and coherent ecosystem of clinical information models, known as openEHR archetypes. Each archetype is designed as a maximal data set for a universal use-case, intended for reuse across various health data sets, known as openEHR templates. To date, only anecdotal evidence has been available to indicate if the hypothesis of archetype reuse across templates is feasible and scalable. As a response to the COVID-19 pandemic, between February and July 2020, 7 openEHR templates were independently created to represent COVID-19–related data sets for symptom screening, confirmed infection reporting, clinical decision support, and research. Each of the templates prioritized reuse of existing use-case agnostic archetypes found in openEHR International's online Clinical Knowledge Manager tool as much as possible. This study is the first opportunity to investigate archetype reuse within a range of diverse, multilingual openEHR templates.

Objective: This study aims to investigate the use and reuse of openEHR archetypes across the 7 openEHR templates as an initial investigation about the reuse of information models across data sets used for a variety of clinical purposes.

Methods: Analysis of both the number of occurrences of archetypes and patterns of occurrence within 7 discrete templates was carried out at the archetype or clinical concept level.

Results: Across all 7 templates collectively, 203 instances of 58 unique archetypes were used. The most frequently used archetype occurred 24 times across 4 of the 7 templates. Total data points per template ranged from 40 to 179. Archetype instances per template ranged from 10 to 62. Unique archetype occurrences ranged from 10 to 28. Existing archetype reuse of use-case agnostic archetypes ranged from 40% to 90%. Total reuse of use-case agnostic archetypes ranged from 40% to 100%.

Conclusions: Investigation of the amount of archetype reuse across the 7 openEHR templates in this initial study has demonstrated significant reuse of archetypes, even across unanticipated, novel modeling challenges and multilingual deployments. While the trigger for the development of each of these templates was the COVID-19 pandemic, the templates represented a variety of types of data sets: symptom screening, infection report, clinical decision support for diagnosis and treatment, and secondary use or research. The findings support the openEHR hypothesis that it is possible to create a shared, public library of standards-based, vendor-neutral clinical information models that can be reused across a diverse range of health data sets.

KEYWORDS

Introduction

Background

Despite electronic health records being in existence for over 50 years, our ability to exchange health data remains frustratingly limited. Semantic interoperability, as defined by the Healthcare Information and Management Systems Society [1], "provides for common underlying models and codification of the data including the use of data elements with standardised definitions from publicly available value sets and coding vocabularies, providing shared understanding and meaning to the user."

We have many long-established terminologies from which we can draw coded value sets, such as SNOMED-CT (SNOMED Clinical Terms) [2], LOINC (Logical Observation Identifiers Names and Codes) [3], or ICNP (International Classification for Nursing Practice) [4]. Commonly used clinical content standards, and the information models that underpin them, are primarily related to health data exchange, and so are usually document- or message-focused [5]. In contrast to this, there have been two primary efforts to develop standards-based atomic clinical information models—the HL7 [6] Clinical Information Modelling Initiative (CIMI) [7] and the openEHR International [8] Clinical Models program [9]. Each of these groups aims to establish an open and shared library of standards-based, vendor-neutral, and use-case agnostic information models representing clinical concepts.

The vision of creating a public library of information models that potentially hold the whole scope, breadth, depth, and range of the health care domain is, at the very least, rather daunting. It is effectively seeking to establish a governed, coordinated, and coherent ecosystem of health data definitions. The goal is to develop each information model once and reuse them across various health data sets, potentially including those for data exchange, health record persistence, data registries, population health, and research. Due to the novelty of this approach, there is only anecdotal evidence so far on its feasibility.

In response to the COVID-19 pandemic, between February and July 2020, 7 openEHR templates were independently created to represent COVID-19–related data sets for symptom screening, confirmed infection reporting, clinical decision support, and research. Each of the templates prioritized reuse of existing use-case agnostic archetypes found in openEHR International's online Clinical Knowledge Manager [10] (CKM) tool as much as possible.

This case study aims to investigate the use and reuse of openEHR archetypes across the 7 openEHR templates as an initial investigation on the reuse of information models across data sets used for a variety of clinical purposes.

The openEHR Approach: Archetypes and Templates

Since 2008, the openEHR Clinical Models program has developed a comprehensive and collaborative methodology to develop clinical information models known as openEHR archetypes [11]. It has gradually developed an extensive library of high-quality, multilingual, and use-case agnostic archetypes that can then be aggregated, constrained, and reused in implementable data sets known as openEHR templates.

An openEHR archetype is a computable specification for a single clinical concept, based on the ISO 13606-2 Archetype interchange specification [12]. The archetypes represent clinical knowledge in a consistent, formal, computable format, independent of any software application or technical implementation. Combined with terminology, they provide a standardized and consistent way to capture, store, display, exchange, aggregate, and analyze health data.

The openEHR approach is unique in that an archetype design strategically aims for a notional maximal data set of relevant data elements with a use-case agnostic mindset to support all possible use—the universal use-case. Achieving a complete maximal data set or inclusion of all use cases is impossible to determine, except perhaps with the benefit of hindsight; however, it is the philosophical avoidance of a minimum data set approach that is critical. Best practice in archetype design aims for each archetype to include all data elements useful to express all attributes about the clinical concept, associated metadata describing the concept, use and misuse, and translations from the original authoring language.

Templates represent a specific data set, comprising one or more archetypes that are constrained to accurately match the data set requirements for a particular clinical use case, health domain, profession, or geographical location. The number of archetypes used in a template reflects the required scope of content and level of detail. Some simple templates representing a laboratory test report may comprise only a single archetype. Theoretically, there is no upper limit to the number of archetypes included in a single template. In practice, a consultation note for a first antenatal visit could comprise data elements from 50 or more archetypes to embrace the diversity and detail of clinical information required for an initial pregnancy assessment.

Principles of template development methodology strongly encourage reuse of existing published openEHR archetypes where available, customize existing archetypes to fit the clinical use case, and develop new archetypes only in situations where no previous archetype exists.

The two-level modeling approach described in the openEHR Archetype formalism [13]—defining and standardizing archetypes first, followed by combining and constraining them to create clinical templates—is unique to the openEHR approach. The rigorously governed, published archetypes held in the CKM provide a robust clinical knowledge foundation. Simultaneously, the templates enable modelers to represent diverse and complex real-world clinical information in standardized data sets.

The underlying crowdsourcing approach highlights openness, transparency, and accountability to the openEHR community. The CKM is a critical enabler: an online hub providing a shared library of archetypes and templates; a collaboration portal receiving contributions of models and expertise from the international member community; and a governance tool to manage clinical content publication, language translation, and artifact versioning.

The CKM volunteer community has over 2500 registered users from 103 countries—comprising clinicians, informaticians,

software engineers, terminologists, academics, students, and consumers. There are 535 governed archetypes in CKM: 115 of these archetypes have completed peer review and have been published; 26 are currently undergoing a peer-review process of the content; the remainder are candidates for future publication. With an average of 15 data points per archetype, the CKM library equates to more than 8000 data points.

English is the original language of the international CKM, but each archetype can be multilingual. Translation of archetypes is a significant activity by community volunteers. Currently, CKM contains archetypes in 29 languages, with the most common translations into Norwegian Bokmål, German, and Portuguese (Brazil).

Case Study

In response to the COVID-19 pandemic, several implementers within the openEHR community openly shared their COVID-19–related templates in CKM. It started with one vendor and grew organically into a grassroots, community-driven collaboration. Three phases of template development were identified.

Phase 1

In late February, a major Norwegian hospital vendor recognized the need to develop and deploy new software tools in their clinical system to equip clinicians to monitor and report on COVID-19 cases within their hospitals. Within 10 days, two openEHR templates were created and deployed in English and Norwegian Bokmål for use in clinical systems [14,15] in Norway, Slovenia, and the United Kingdom. The templates were uploaded to a CKM COVID-19 public incubator [16] in early March 2020 under a CC-BY license:

- Template 1: Suspected COVID-19 Risk Assessment [17] – based on guidance from advice given by the World Health Organization (WHO) and public health authorities in the United Kingdom, Slovenia, New Zealand, and Norway.
- Template 2: Confirmed COVID-19 Infection Report [18] – replicates a WHO case report form [19].

Due to the rapid deployment priorities imposed by the pandemic, the primary author of templates 1 and 2 developed the templates by reuse of as many existing archetypes as possible and opted to create new COVID-19–specific archetypes to represent the remaining data points (Ian McNicoll, McChB, personal communication). It was a reasonable and pragmatic design decision, with well-understood consequences—effectively a compromise between strategic reuse design principles and speed of modeling.

Phase 2

Soon after templates 1 and 2 were uploaded, CKM editors reviewed the COVID-19–specific archetypes. The editors analyzed the requirements from phase 1 and identified the missing archetype concepts in the CKM library. In direct response to the gap analysis, 11 new content-equivalent, use-case agnostic archetypes were created, covering screening questionnaires, travel history, and infectious disease exposure.

Both phase 1 templates were revised and updated using only use-case agnostic archetypes and uploaded to the CKM COVID-19 project [20]:

- Template 3: Suspected COVID-19 Risk Assessment2 [21] – a revised version of template 1
- Template 4: Confirmed COVID-19 Infection Report2 [22] – a revised version of template 2

Also, the modelers created 1:1 mappings [23,24] of all data points from templates 1 to 3 and templates 2 to 4 with a 98% success rate, providing a future migration path should the clinical systems using templates 1 and 2 choose to upgrade to the revised templates and use-case agnostic archetypes.

Phase 3

Three more templates were developed in the weeks and months that followed, and these provided an opportunity to test if the new phase 2 archetypes were fit for purpose and able to represent the requirements for other COVID-19–related data sets.

A Chinese university developed an entirely new template, representing the official Chinese Guidelines for Diagnosis and Treatment Guideline of COVID-19 (7th Edition) [25]. This template was used as the basis for a decision support system implemented within a Chinese hospital system in Wuhan, China, deployed in Chinese [26]. It was uploaded to CKM in early April 2020:

- Template 5: COVID-19 Pneumonia Diagnosis and Treatment (7th Edition) [27]

An Italian health software vendor adapted template 3 for a COVID-19 risk screening application with a nephrology focus. First-time modelers developed it for deployment in Italian within Brotzu Hospital, Cagliari, Sardinia [28], and uploaded it to CKM in late April 2020:

- Template 6: Suspected COVID-19 Risk Assessment Nephrology [29]

The Fast Healthcare Interoperability Resources Implementation Guide (FHIR IG) for the German Corona Consensus Data Set (GECCO) [30] supporting COVID-19 research was released in July 2020. In parallel, an openEHR template was developed to replicate the GECCO data set. It was uploaded to CKM in late July 2020:

- Template 7: German Corona Consensus Data Set (GECCO) [31]

Methods

Analysis of both the number of occurrences of archetypes and patterns of occurrence within 7 discrete openEHR templates was carried out at the archetype or clinical concept level.

Reuse per Template

The 7 templates were analyzed in terms of:

- Total data points: the total number of data elements or fields represented within a template, which gives an impression

of the level of complexity or level of detail within the template;

- Archetype instances: the total number of archetypes used within a template, including reuse;
- Unique archetype occurrences: the total number of archetypes used within a template, excluding any reuse or repetition of archetypes;
- Existing archetype reuse %: the number of use-case agnostic archetypes that existed in CKM before phase 1 that were used in the template, expressed as a percentage of the total number of archetype instances in the template;
- New archetype reuse %: the number of use-case agnostic archetypes that were created during phase 2, used in the template and uploaded to the CKM pool, expressed as a percentage of the total number of archetype instances in the template;
- COVID-specific archetype use %: the number of COVID-19–specific archetypes created during phase 1 that were used in the template, expressed as a percentage of the total number of archetype instances in the template.

Reuse per Archetype

The 7 templates were analyzed in terms of:

- Archetype reuse: the number of times an archetype was used across all templates;
- Template count: the number of templates that contained at least one occurrence of an archetype.

Results

Reuse per Template

The results in Table 1 focus on the overall archetype composition of each template. The three archetype categories in Table 1 are defined as:

- Existing archetypes: archetypes that had been authored before the COVID-19 template was developed and available in CKM;
- New archetypes: archetypes authored as use-case agnostic archetypes during phase 2;
- Phase 1 archetypes: COVID-19–specific archetypes authored during phase 1.

Total data points per template ranged from 40 to 179. The template with the largest number of data points was template 5, the Chinese COVID-19 guideline data set. Archetype instances per template ranged from 10 to 62. The template with the largest number of archetype instances was template 5. Unique archetype occurrences ranged from 10 to 28. The template with the largest number of unique archetypes was template 7, the German GECCO data set. Total reuse of use-case agnostic archetypes ranged from 40% to 100%. Existing archetype reuse of use-case agnostic archetypes ranged from 40% to 90%, and new archetype reuse of use-case agnostic archetypes ranged from 0% to 43%.

COVID-19–specific archetypes created for novel clinical concepts were used in templates 1 and 2. New use-case agnostic archetypes replacing the COVID-19–specific archetypes were used within templates 3, 4, 5, 6, and 7. Template 6 used a combination of ungoverned and new archetypes.

The number of data points per template can be considered a proxy for the level of detail in the template. The number of unique archetypes per template reflects the diversity of clinical content in the clinical requirements and may be considered a proxy for the level of complexity in the template.

Table 1. Reuse per template.

Variable	Template						
	1	2	3	4	5	6	7
Total data points, n	60	88	40	102	179	129	124
Archetype instances, n	15	21	16	26	62	10	53
Unique archetype occurrences, n	10	14	16	16	17	10	28
Total reuse (%)	40	52	100	100	100	80	100
Existing archetype reuse, n/N (%)	6/15 (40)	11/21 (52)	13/16 (81)	16/26 (62)	56/62 (90)	5/10 (50)	30/53 (57)
New archetype reuse, n/N (%)	0/15 (0)	0/21 (0)	3/16 (19)	10/26 (38)	6/62 (10)	3/10 (30)	23/53 (43)
COVID-19–specific archetype use, n/N (%)	9/15 (60)	10/21 (48)	0/16 (0)	0/26 (0)	0/62 (0)	2/10 (20)	0/53 (0)

Reuse per Archetype

The results in Table 2 focus on archetype concept reuse within templates by examining how many times each archetype occurs in each template. For example, the first archetype "Laboratory test result" is a published archetype and was used twice in template 2, twice in template 4, 15 times in template 5, and 5 times in template 7, for a total of 24 instances of reuse across 4 templates. Only clinical archetype use within the templates was analyzed.

Table 2. Reuse per archetype (total number of archetype instances=203).

Archetype concept name	Publication status[a]	Template							Archetype reuse	Template count
		1	2	3	4	5	6	7		
Existing archetypes										
Laboratory test result	P		2		2	15		5	*24* [b]	*4*
Laboratory analyte result	P					15		4	*19*	*2*
Specimen	P					15			*15*	*1*
Address	D		3	1	3				*7*	*3*
Problem/diagnosis	P	1		1	2			3	*7*	*4*
Health risk assessment	P			1	4		1		*6*	*3*
Story/History	P	1	2	1	1		1		*6*	*5*
Body temperature	P	1		1		1		1	*4*	*4*
Encounter	P	1		1			1		*3*	*3*
Facility	D		1	1	1				*3*	*3*
Occupation record	P		1	1	1				*3*	*3*
Occupation summary	P		1	1	1				*3*	*3*
Imaging test result	D					2		1	*3*	*2*
Problem/diagnosis qualifier	P	1		1				1	*3*	*3*
Dwelling	D			1			1		*2*	*2*
Living arrangement	D			1			1		*2*	*2*
Imaging finding	D					2			*2*	*1*
Report	P		1		1				*2*	*2*
Service request	P	1		1					*2*	*2*
Inspired oxygen	P					1		1	*2*	*2*
Pulse oximetry	P					1		1	*2*	*2*
Respiration	P					1		1	*2*	*2*
Symptom/sign	P							1	*1*	*1*
Age	D					1			*1*	*1*
Pulse	P							1	*1*	*1*
Blood pressure	P							1	*1*	*1*
Differential diagnoses	R					1			*1*	*1*
SOFA[c] score	D							1	*1*	*1*
PaO2/FiO2 ratio	P					1			*1*	*1*
Tobacco smoking summary	P							1	*1*	*1*
Exclusion - specific	P							1	*1*	*1*
Gender	P							1	*1*	*1*
Clinical frailty score	P							1	*1*	*1*
Body weight	P							1	*1*	*1*
Body height	P							1	*1*	*1*
Resuscitation status –United Kingdom	D							1	*1*	*1*
Medication summary	D							1	*1*	*1*
New archetypes										
Condition screening questionnaire	D				1	1		12	*14*	*3*

Archetype concept name	Publication status[a]	Template							Archetype reuse	Template count
		1	2	3	4	5	6	7		
Symptom/sign screening questionnaire	D			1	2	1	1	1	*6*	*5*
Management screening questionnaire	D				2	1		3	*6*	*3*
Travel event	D			1	1		1	1	*4*	*4*
Medication screening questionnaire	D							4	*4*	*1*
Overcrowding screening	D			1		1			*2*	*2*
Exposure assessment questionnaire	D					1		1	*2*	*2*
Procedure screening questionnaire	D				2				*2*	*1*
Therapeutic order	D					2			*2*	*1*
Episode of care	D				1			1	*2*	*2*
Infectious exposure investigation	D				1				*1*	*1*
Phase 1 COVID-19–specific archetypes										
Symptom/sign – COVID-19	X	6	2						*8*	*2*
Admission	X		2						*2*	*1*
Outbreak exposure	X	1	1						*2*	*2*
Procedure summary	X		2						*2*	*1*
Travel trip history	X	1	1						*2*	*2*
Comorbidity summary	X		1						*1*	*1*
COVID outcomes	X		1						*1*	*1*
Fever (*Febbre* in Italian)	X					1			*1*	*1*
Health risk assessment –COVID-19	X	1							*1*	*1*
Social summary COVID-19	X					1			*1*	*1*

[a]The publication status key is: P=content is published; R=content is undergoing peer review; D=draft candidate; X=ungoverned, specific use-case only.
[b]Important values are italicized.
[c]SOFA: sequential organ failure assessment score.

Across all 7 templates collectively, 203 instances of 58 unique archetypes were used.

There were 48 existing and new use-case agnostic archetypes, of which 26 had completed the peer-review process, and the content published; 1 is currently undergoing the peer-review process; the remaining 21 are draft candidates.

The "Laboratory test result" archetype was the most reused archetype; it occurred 24 times across 4 templates. Reuse across templates reflects the commonality of content, despite different design intents for the templates. The existing "Story/History" and the new "Symptom/sign screening questionnaire" archetypes were reused across the largest number of templates—within 5 out of the 7 templates each.

Many of the existing archetypes were only used once within the context of these 7 templates. These archetypes had been authored for use-case agnostic use before the development of the COVID-19 data sets, so any reuse within these COVID-19 templates demonstrates reusability across both COVID-19 and non–COVID-19 use cases.

Not all archetypes were used in each template, reflecting the diversity of content requirements across the 7 templates.

Discussion

Principal Findings

Before February 2020, the focus of openEHR International's CKM was on creating a library of shared archetypes. Templates had been uploaded to CKM, most commonly to demonstrate modeling patterns or to provide exemplars for common types of data sets. Any estimates of reuse of archetypes across templates had been wholly anecdotal, communicated directly by experienced modelers, and ranged from 60% to 90% reuse.

The onset of the COVID-19 pandemic triggered a collaborative openEHR community effort to fast track both archetype and template development, with CKM used as a coordinating hub. The 7 templates uploaded during this time to CKM have provided the first opportunity for a formal analysis of reuse.

Public sharing of the initial templates, templates 1 and 2, included 8 COVID-19–specific archetypes that were necessitated by the novel content combined with rapid implementation deadlines, resulting in relatively low reuse (ie, 40% and 52%, respectively). This was a reasonable and pragmatic modeling decision in the circumstances but diverged from the

recommended design philosophy aiming for use-case agnostic archetypes that usually take more time to develop.

Soon after, the CKM editors redesigned the 8 COVID-19–specific archetypes as 11 new use-case agnostic revisions—conceptually equivalent but intentionally designed to allow for broader reuse—and uploaded them as additions to the CKM library. The clinical concepts modeled in the new archetypes included a range of clinical screening question/answer pairs, as well as models for travel history and a risk assessment about exposure to infectious agents. Revised versions of the initial templates, with 100% reuse, were uploaded as templates 3 and 4, along with associated data mappings.

Modeling of questionnaire archetypes had been attempted unsuccessfully in the past, but without success [32]. Driven by the new COVID-19 screening requirements, modelers revisited the challenge of questionnaire modeling. Subsequently, they developed a family of screening questionnaire archetypes that were use-case agnostic and based on an underlying shared pattern, covering the screening for symptoms and signs, conditions, procedures, management and treatment, medication use, and exposure to agents. They were uploaded to a dedicated project in CKM [33] and made available for broader community reuse.

Phase 3 template development provided an opportunity to test and confirm the reuse potential for the new archetypes in additional clinical data sets.

Template 5 represented the official Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment and was implemented as the foundation for a decision support application. This data set was the most extensive and most detailed in terms of both the number of data points and the number of archetype instances. Laboratory and imaging test results triggered system-generated advice about diagnosis and treatment, resulting in high reuse of the "Laboratory test result," "Laboratory analyte result," and "Specimen" archetypes. This template achieved 100% reuse of 17 unique archetypes drawn from the "existing" and "new" archetype pools. The archetypes were all translated into Chinese and uploaded to a Chinese equivalent of the CKM tool, known as the Healthcare Modelling Collaboration tool [34].

Template 6 represented Suspected COVID-19 Risk Assessment data within a nephrology context. It was based on template 3, including the screening questionnaire archetypes but reuse was reduced to 80% due to the inclusion of the "Fever" and "Social summary" archetypes intended to meet local data requirements.

Template 7 was created after communication with the authors of the FHIR IG for the GECCO. It was developed to investigate if the clinical content of a data set explicitly developed for implementation in FHIR could also be represented using openEHR archetypes. The resulting template contains the largest number of unique archetypes, which strongly suggests that this template was the most complex of the 7 templates. It was developed in 4 hours and resulted in 100% archetype reuse of 28 unique archetypes drawn from the "existing" and "new" archetype pools. Creation of the template first involved investigation and analysis of the FHIR IG to identify the clinical

requirements and archetypes required, followed by aggregation and constraint of each archetype to match the precise requirements of the FHIR data set. Terminology value sets were not included in the modeling as it was assumed that the same value sets in the FHIR IG would be applicable in the openEHR template.

While there is considerable diversity in purpose or intent across the 7 templates, the level of archetype reuse is a clear indication of the level of commonality in the clinical concepts that underpin each data set. In addition, even though the focus and level of detail for each template varied, the shared data models underpinning each template ensured consistency of data across all of them.

It is also important to note the maturity of the archetypes used—70% (26/37) of the "existing" archetypes have completed the content peer-review process and have been published, which may be considered a proxy for data quality. Further investigations about the qualitative and quantitative assessment of archetype quality should be undertaken—firstly to assess each archetype, but also as a proxy for broader data set quality.

In building a template for each new data set, the amount of reuse depends on the similarity of its clinical concepts with archetypes created for inclusion in prior data sets. It is not so much the purpose, level of detail, or complexity of the data set that influences reuse, but rather the commonality of the component clinical concepts that determine which archetypes are required. In practice, each new template developed leverages all prior work that has shaped each existing archetype in the CKM library and, as illustrated by the development of new archetypes for templates 3 and 4, often extends the library collection. The design approach of archetypes as maximal data sets and universal use case for each concept supports the representation of a variety of levels of detail required in data sets. New clinical requirements are added by extending existing archetypes or creating new archetypes for novel concepts. Over time we can expect the number of archetypes to continue to grow and archetype quality enhanced with increasing levels of detail and refinements from the peer-review process. In this context, it is not unreasonable to expect future archetype reuse to remain at similar rates to those demonstrated in this set of COVID-19–related templates.

The 11 new archetypes in phase 2 were strategically designed as draft candidates: aiming for an inclusive, maximal data set about a single clinical concept; intended for a universal use case; discrete in scope, without any overlap with other archetypes.

The current CKM archetype library comprises a range of archetypes used in prior work. Each new, use-case agnostic archetype developed as part of the creation of the COVID-19–related templates added to CKM will be available for reuse in future modeling efforts. In this way, the CKM library will continue to grow, underpinned by technical and editorial governance processes to ensure coordination and coherence of the archetype library.

In this study, we have observed how the collection of archetypes listed in Table 2, a subset of the CKM library, has provided a

focused ecosystem of coordinated and coherent information models to underpin each of the 7 data sets. With the whole CKM comprising 500+ archetypes and 8000+ data elements, it becomes more plausible to imagine the potential for this more extensive library of standardized, coordinated, and coherent information models to be able to represent a broader and more diverse range of data sets. In addition, as in the case of the development of template 7, if reuse of archetypes enables the creation of a template comprising 124 data points within 4 hours, the potential time efficiencies gained through archetype reuse is also worthy of further investigation to determine if this is more broadly applicable.

Conclusion

Investigation of the amount of archetype reuse across the 7 openEHR templates in this initial study has demonstrated significant reuse of archetypes, even across unanticipated, novel modeling challenges and multilingual deployments. While the trigger for the development of each of these templates was the COVID-19 pandemic, the templates represented a variety of types of data sets— symptom screening, infection report, clinical decision support for diagnosis and treatment, and secondary use or research.

The findings support the openEHR hypothesis that it is possible to create a shared, public library of standards-based, vendor-neutral clinical information models that can be reused across a diverse range of health data sets.

Further investigation is strongly recommended to evaluate:

- The realistic extent and scope of a shared library of information models, including the limitations and barriers. Is it plausible to create a single universal health language, or would it be more feasible to develop libraries for specific purposes?
- Clinical knowledge governance requirements for a library of shared information models;
- The measurement of the quality of individual information models;
- The impact on data set quality if based on a foundation of high-quality information models;
- Time and cost efficiencies of creating data sets from a shared library of information models;
- The impact on health data interoperability if shared information models are used as the basis of data exchange directly between clinical systems, in different contexts, and for various purposes;
- The impact on clinical safety when information models are shared and the need for data transformation or mapping is reduced or eliminated;
- The impact on secondary use of data and research if shared information models are used, supporting safe and accurate aggregation and analysis of health data.

Conflicts of Interest

HL is the colead for openEHR International's Clinical Models program.

References

1. Interoperability in Healthcare? What is Interoperability? HIMSS. 2020. URL: https://www.himss.org/resources/interoperability-healthcare#what_is_interoperability [accessed 2020-09-14]
2. SNOMED International. 2020. URL: http://www.snomed.org/ [accessed 2020-09-14]
3. LOINC. 2020. URL: https://loinc.org/ [accessed 2020-09-14]
4. International Council of Nurses. 2020. URL: https://www.icn.ch/what-we-do/projects/ehealth-icnptm [accessed 2020-09-14]
5. Interoperability in Healthcare? Standards Development. HIMSS. 2020. URL: https://www.himss.org/resources/interoperability-healthcare#interoperability_standards [accessed 2020-09-14]
6. Health Level Seven International. 2020. URL: https://www.hl7.org/ [accessed 2020-09-14]
7. Clinical Information Modeling Initiative. Health Level Seven International. 2020. URL: https://www.hl7.org/Special/Committees/cimi/ [accessed 2020-09-14]
8. openEHR International. 2020. URL: https://openehr.org/ [accessed 2020-09-14]
9. Clinical Models program. openEHR International. 2020. URL: http://www.openehr.org/programs/clinicalmodels/ [accessed 2020-09-14]
10. Clinical Knowledge Manager. openEHR International. 2020. URL: http://www.openehr.org/ckm/ [accessed 2020-09-14]
11. Leslie H, Ljosland Bakke S. Peer Review of Clinical Information Models: A Web 2.0 Crowdsourced Approach. Stud Health Technol Inform 2017;245:905-909. [Medline: 29295231]
12. ISO 13606-2:2019(en) Health informatics — Electronic health record communication — Part 2: Archetype interchange specification. International Organisation for Standardisation. 2019. URL: https://www.iso.org/obp/ui/#iso:std:iso:13606:-2:ed-2:v1:en [accessed 2020-09-14]
13. Archetype Formalism Overview. openEHR International. 2020. URL: https://specifications.openehr.org/releases/AM/latest/Overview.html#_archetype_formalism_overview [accessed 2020-09-14]
14. openEHR Community Rises to the Challenge of Coronavirus. Open Health News. 2020 Mar 11. URL: http://www.openhealthnews.com/content/openehr-community-rises-challenge-coronavirus> [accessed 2020-09-14]
15. McNicoll I. openEHR Discourse discussion forums. openEHR COVID-19 Project. 2020 Mar 10. URL: https://discourse.openehr.org/t/openehr-covid-19-project/448 [accessed 2020-09-14]

XSL•FO

RenderX

16. COVID-19 Incubator. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/projects/1013.30.80 [accessed 2020-09-14]

17. Suspected COVID-19 risk assessment, draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.267 [accessed 2020-09-14]

18. Confirmed COVID-19 infection report, draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.271 [accessed 2020-09-14]

19. Revised Case Report Form For Confirmed Novel Coronavirus COVID-19. World Health Organization. 2020 Feb 27. URL: https://apps.who.int/iris/bitstream/handle/10665/331234/WHO-2019-nCoV-SurveillanceCRF-2020.2-eng.pdf [accessed 2020-09-14]

20. COVID-19 Project. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/projects/1013.30.81 [accessed 2020-09-14]

21. Suspected COVID-19 risk assessment2, draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.286 [accessed 2020-09-14]

22. Confirmed COVID-19 infection report2, draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.282 [accessed 2020-09-14]

23. Mapping spreadsheet for openEHR-Confirmed Covid-19 infection report.v0, draft template. openEHR Foundation, openEHR Clinical Knowledge Manager. URL: https://ckm.openehr.org/ckm/templates/1013.26.271/resourcecentre [accessed 2020-09-14]

24. Mapping spreadsheet for openEHR-Suspected Covid-19 assessment.v0, draft template. openEHR Foundation, openEHR Clinical Knowledge Manager. URL: https://ckm.openehr.org/ckm/templates/1013.26.267/resourcecentre [accessed 2020-09-14]

25. Chinese Clinical Guidance For COVID-19 Pneumonia Diagnosis And Treatment (7th Edition). China National Health Commission. 2020 Mar 4. URL: http://kjfy.meetingchina.org/msite/news/show/cn/3337.html [accessed 2020-09-14]

26. Li M, Leslie H, Qi B, Nan S, Feng H, Cai H, et al. Development of an openEHR Template for COVID-19 Based on Clinical Guidelines. J Med Internet Res 2020 Jun 10;22(6):e20239 [FREE Full text] [doi: 10.2196/20239] [Medline: 32496207]

27. COVID-19 pneumonia diagnosis and treatment (7th Edition), draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.291 [accessed 2020-09-14]

28. COVID-19 App Di Monitoraggio - Software Demonstration. INPECO Webapp Portal. 2020. URL: https://covid19-webapp.inpecotpm.com/ [accessed 2020-09-14]

29. Suspected COVID-19 Risk assessment nephrology, draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.298 [accessed 2020-09-14]

30. German Corona Consensus Data Set - Implementation Guide. Simplifier.net. 2020. URL: https://simplifier.net/guide/GermanCoronaConsensusDataSet-ImplementationGuide/Covid-19therapy [accessed 2020-09-14]

31. German Corona Consensus Data Set (GECCO), draft template. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/templates/1013.26.372 [accessed 2020-09-14]

32. Leslie H. Blog - The Questionnaire Challenge. Atomica Informatics. 2014 Feb 22. URL: https://www.atomicainformatics.com/archetypical/2014/02/21/the-questionnaire-challenge [accessed 2020-09-14]

33. Screening Questionnaire Project. openEHR Clinical Knowledge Manager. 2020. URL: https://ckm.openehr.org/ckm/projects/1013.30.90 [accessed 2020-09-14]

34. Healthcare Modelling Collaboration. Zhejiang University. URL: http://openehr.org.cn/hmc/ [accessed 2020-09-14]

Abbreviations

CIMI: Clinical Information Modelling Initiative
CKM: Clinical Knowledge Manager
FHIR: Fast Healthcare Interoperability Resources
GECCO: German Corona Consensus Data Set
ICNP: International Classification for Nursing Practice
IG: Implementation Guide
LOINC: Logical Observation Identifiers Names and Codes
SNOMED-CT: SNOMED Clinical Terms
WHO: World Health Organization

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

Exploring the Role of Media Sources on COVID-19–Related Discrimination Experiences and Concerns Among Asian People in the United States: Cross-Sectional Survey Study

Nan Yu[1], PhD; Shuya Pan[2,3], PhD; Chia-chen Yang[4], PhD; Jiun-Yi Tsai[5], PhD

[1]Nicholson School of Communication and Media, University of Central Florida, Orlando, FL, United States

[2]School of Journalism and Communication, Renmin University of China, Beijing, China

[3]Center of Journalism and Social Development, Renmin University of China, Beijing, China

[4]School of Educational Foundations, Leadership and Aviation, Oklahoma State University, Stillwater, OK, United States

[5]School of Communication, Northern Arizona University, Flagstaff, AZ, United States

Corresponding Author:
Shuya Pan, PhD
School of Journalism and Communication
Renmin University of China
59 Zhongguancun St
Beijing, 100082
China
Phone: 86 1082500855
Email: shuya@ruc.edu.cn

Abstract

Background: Media coverage and scholarly research have reported that Asian people who reside in the United States have been the targets of racially motivated incidents during the COVID-19 pandemic.

Objective: This study aimed to examine the types of discrimination and worries experienced by Asians and Asian Americans living in the United States during the pandemic, as well as factors that were associated with everyday discrimination experience and concerns about future discrimination that the Asian community may face.

Methods: A cross-sectional online survey was conducted. A total of 235 people who identified themselves as Asian or Asian American and resided in the United States completed the questionnaire.

Results: Our study suggested that up to a third of Asians surveyed had experienced some type of discrimination. Pooling the responses "very often," "often," and "sometimes," the percentages for each experienced discrimination type ranged between 14%-34%. In total, 49%-58% of respondents expressed concerns about discrimination in the future. The most frequently experienced discrimination types, as indicated by responses "very often" and "often," were "people act as if they think you are dangerous" (25/235, 11%) and "being treated with less courtesy or respect" (24/235, 10%). About 14% (32/235) of individuals reported very often, often, or sometimes being threatened or harassed. In addition, social media use was significantly associated with a higher likelihood of experiencing discrimination (β=.18, P=.01) and having concerns about future episodes of discrimination the community may face (β=.20, P=.005). Use of print media was also positively associated with experiencing discrimination (β=.31, P<.001).

Conclusions: Our study provided important empirical evidence regarding the various types of discrimination Asians residing in the United States experienced or worried about during the COVID-19 pandemic. The relationship between media sources and the perception of racial biases in this group was also identified. We noted the role of social media in reinforcing the perception of discrimination experience and concerns about future discrimination among Asians during this outbreak. Our results indicate several practical implications for public health agencies. To reduce discrimination against Asians during the pandemic, official sources and public health professionals should be cognizant of the possible impacts of stigmatizing cues in media reports on activating racial biases. Furthermore, Asians or Asian Americans could also be informed that using social media to obtain COVID-19 information is associated with an increase in concerns about future discrimination, and thus they may consider approaching this media source with caution.

XSL•FO
RenderX

Introduction

Background

As of September 2020, COVID-19 has infected approximately 6.94 million people in the United States and caused more than 200,000 deaths in the country [1]. Globally, over 31 million people had been infected with COVID-19, making it one of the worst pandemics in human history [1,2]. It is common for people to seek the origin of a pandemic and focus on responsibility and assign blame during an outbreak [3]. Media reports and nonprofit organizations have alerted the Asian community in the United States regarding the rise of anti-Asian hate crimes since the start of this pandemic [4-6]. In late March, the Federal Bureau of Investigation also warned this community about a surge in hate crimes during the pandemic and alerted law enforcement agencies to pay closer attention to bias-motivated crimes [7,8].

Several empirical studies have indicated that Asians residing in the United States are vulnerable to racist behaviors during the outbreak of an infectious disease like COVID-19. Dhanani and Franz [9] conducted a national survey of US adults in March 2020 to study COVID-19–related discriminatory behaviors, including avoiding ordering from restaurants with primarily Asian employees, limiting interactions with Asian customers or coworkers, or intentionally moving away from an Asian individual while in a public place. Their study found that 42% of the 1141 people surveyed engaged in at least one discriminatory behavior toward people of Asian descent [9]. Another study showed that racism against Asian groups, which is associated with the outbreak of an infectious disease, is not a new phenomenon but has happened several times throughout American history [7]. Meanwhile, Asians across the globe were also found to experience xenophobia and stigmatization during the SARS (severe acute respiratory syndrome) outbreak in 2003 [10].

At the same time, according to several news reports, Asians and people of Asian descent who reside in the United States had been targets of blame, partly because the virus was first reported in Wuhan, China [11,12]. In addition, US President Donald Trump and other political officials have publicly referred to COVID-19 as "Wuhan virus," "Chinese virus," and "Kung Flu," causing a surge in racist behavior against Asians [5,10,13]. The Asian Pacific Policy and Planning Council received approximately 1500 reports on anti-Asian hate incidents from 45 states during a 4-week period from March to April 2020 [6]. Although there is no existing research directly comparing the prevalence of anti-Asian incidents before and after the COVID-19 outbreak, both academic research and news reports have shown that the Asian community experienced a high level of verbal harassment, shunning, physical assaults, workplace discrimination, being barred from transportation, and being turned away by businesses [5-7,12,13]. The Center of Public Integrity [5] suggested that the national mental health support organization Crisis Text Line [14] received significantly more texts from people of Asian descent after President Trump called the coronavirus "Chinese virus" in late March 2020. During the same period, Reny and Barreto [15] found that Google searches for "Chinese virus" and "Kung Flu" increased dramatically.

Given the current social environment, it is meaningful to study the discrimination experienced by people of Asian descent living in the United States during the COVID-19 pandemic. Recently, scholars have studied relevant topics in the COVID-19 context, such as the impact of discrimination against Asians on mental health [10]; social media usage and anti-Asian sentiments among Caucasians [13]; and anti-Asian attitudes and behaviors in the United States [9].

The rise of social media has enabled it to powerfully influence people's racism-related beliefs and actions. In a recent study, Davidson and Farquhar [16] showed that social media can serve as a major news source and contribute to prejudicial attitudes toward immigrants, refugees, and transgender people. Specifically relating to COVID-19, Croucher et al [13] found that users who perceived information on social media to be accurate and fair were more likely to believe that Chinese Americans could pose realistic threats (eg, threats to physical and material well-being) and symbolic threats (eg, threats to morals, values, and beliefs). Additionally, Ziems et al [17], in a preprint publication, revealed that the presence of anti-Asian hate speech on Twitter was more prevalent than counter-hate messages during the COVID-19 crisis. For Asians living in the United States, being exposed to large amounts of unscreened racist messages on social media could greatly affect their perceptions of discrimination. In fact, a recent study has shown that engagement with social media during the pandemic was associated with worry about discrimination among the Asian individuals surveyed [18].

According to cultivation theory [19], traditional media use could effectively construct a social reality as portrayed by the media sources among viewers. Being exposed to increasing racism-related media content during the COVID-19 pandemic, Asians may form the belief that anti-Asian racism is a salient concern in the United States, leading them to become more sensitive to the discrimination they experience in their daily lives as well as become more worried about the discrimination they may encounter in the future.

Scholars have noted that media exposure can contribute to beliefs and behaviors associated with social biases [16,20,21]. Research also showed that the media can reinforce the marginalization of ethnic minorities and stigmatized groups by portraying them as primary health threats to others [22]. However, previous research mainly documented how racism-related mainstream media coverage (eg, newspapers and TV) could cultivate dominant groups' prejudicial attitudes toward racial minorities. Little is known about how exposure to different media sources could relate to the prejudiced/discriminated groups' experiences and perceptions, which arguably reflect beliefs associated with social biases from

the opposite perspective (ie, the perspective from the prejudiced and discriminated).

To advance these lines of research and create tailored interventions to promote the well-being of Asian people, our study aimed to examine the types of discrimination Asians in the United States have experienced and worried about during the pandemic, as well as how exposure to various traditional media and social media sources could influence experienced and expected future discrimination in this group during the COVID-19 pandemic. Our study is one of the first attempts to focus on Asians residing in the United States during the pandemic and examine factors influencing their experiences of discrimination.

Aims of This Study

Racism refers to a type of ideology in which certain racial groups are viewed as superior to others [23], and racial biases are negative attitudes and beliefs of outgroup members [23-25]. Discrimination is described as unfair treatment and harassment caused by racial biases [24]. In this study, we focused on everyday discrimination, which refers to discriminative events that happen repeatedly in a variety of contexts, such as being treated disrespectfully, being shunned, being offered poor service, or being verbally or physically harassed [26]. The Everyday Discrimination Scale [24] adopted in this study is among the most widely used instruments to evaluate this type of discrimination [26].

We distinguished between two dimensions of discrimination: (1) experiences of discrimination and (2) concerns about future discrimination. Experiences of discrimination refer to people's perceptions of the various forms of anti-Asian incidents encountered during COVID-19, whereas concerns related to future discrimination refers to the extent to which people worried about discrimination that might happen in the future. Our first research question focused on the prevalence of experiences related to discrimination and concerns about future discrimination among Asians who resided in the United States during the pandemic. Our second research question focused on examining the relationship between media sources and Asian participants' experience of discrimination and concerns about future discrimination.

Methods

Sampling

Upon institutional review board (IRB) approval, a cross-sectional online survey was generated on Qualtrics and then distributed to eligible participants via Amazon Mechanical Turk (MTurk). MTurk subject pools have revealed proven advantages in recruiting participants during ongoing social events [27-29]. We posted our Qualtrics survey link on the MTurk platform, through which participants can join the study. MTurk helped us reach eligible participants, and responses from participants were stored in Qualtrics.

Eligible participants included individuals aged ≥18 years who identified as Asian and lived in the United States when the data were collected. To ensure that participants paid sufficient attention during survey completion and to exclude bots and

responses from server farms, we included an attention check question in the survey. Specifically, the question, which was placed randomly within the questionnaire, asked participants to select a designated answer (eg, please select "never" here to show that you are paying attention); those who failed to select the designated answer were excluded from the study.

Data collection was completed in the first 2 weeks of May 2020. At that time, there was a significant number of confirmed COVID-19 cases in the United States, but some states had loosened up the stay-at-home orders. A total of 235 people who identified themselves as Asians or Asian Americans and resided in the United States completed the questionnaire. The sample size for this study was determined by an a priori estimate utilizing G*Power, version 3.1 (Heinrich-Heine-Universität Düsseldorf) [30]. With the regression effect size (R^2) of 0.15, 95% power, and 14 predictors, the minimum sample size was calculated to be 153. In total, 202 participants were included in the regression analyses.

Independent Variable: Media Source

The respondents were asked to indicate how often they obtained news regarding COVID-19 from different media sources. The items were rated on a 4-point scale (1=never to 4=often). Responses indicated media usage as follows: (1) newspapers or magazines (mean 1.69, SD 0.92), (2) radio (mean 1.85, SD 0.97), (3) TV (mean 2.50, SD 1.01), (4) social media (mean 3.22, SD 0.92), and (5) news websites or apps (mean 3.31, SD 0.88).

Dependent Measures

Experience of Discrimination

Experience of discrimination was assessed by asking participants how often they encountered different forms of unfair treatment since the COVID-19 outbreak. The measurement items were adapted from the Everyday Discrimination Scale [24], which is a widely used instrument to evaluate discrimination [26]. The items were rated on a 5-point scale (1=never to 5=very often). This item was worded as follows: "Since the COVID-19 outbreak, in your day-to-day life, how often have any of the following things happened to you: (1) you are treated with less courtesy or respect than other people; (2) people act as if they think you are dangerous; (3) people act as if they are afraid of you; (4) you receive poorer service than other people; (5) you are threatened or harassed." Responses were averaged to indicate the overall experiences of discrimination participants perceived (Cronbach α=.94, mean 1.93, SD 0.98).

Concerns About Future Discrimination

This item measures the degree to which individuals worry about discrimination they may experience in the future. The same set of items in the Everyday Discrimination Scale [24] were used by changing them to future tenses (eg, "how often do you worry that you will be treated with less courtesy or respect than other people?"). Items were averaged to show the overall level of concerns about future discrimination (Cronbach α=.96, mean 2.61, SD 1.18).

Data Analysis

Data were analyzed with SPSS Statistics 25 (IBM Corp). Descriptive statistics were used to investigate the prevalence of different types of discriminatory experiences Asians have experienced or expected to encounter in the future. Two hierarchical regression analyses were used to investigate factors that were associated with experienced everyday discrimination and concerns about future discrimination, respectively. Model 1 included demographic variables such as age, years living in the United States, sex, being Chinese, being a visitor (ie, not a US citizen or permanent resident), education level, employment status, income, and political orientation. Model 2 included the variables in Model 1 as well as different media sources. Some demographic variables were recoded for the regression analyses (sex: male=1, female=0; being Chinese: yes=1, no=0; visitor: visitor=1; US citizen and permanent resident=0; and employment status: employed=1; no=0). Age and years living in the United States were measured as continuous variables. Education level ranged from 1=<high school degree to 7=doctoral degree; income ranged from 1=<$10,000 to 12=≥$150,000; and political orientation varied from 1=very liberal to 7=very conservative. Assumptions, including normality, homoscedasticity, and multicollinearity, were checked before regression tests were performed. All assumptions were met. For the multicollinearity check, variance inflation factor values ranged between 1-2 and far below 10.

Results

Participant Demographics

In total, 235 individuals completed the survey. The participants' age ranged from 18 to 73 years (mean 32.87, SD 10.95 years).

Participants from 35 out of 50 states in the United States comprised the sample and had lived in the country for an average of 23.56 years (SD 12.95). Approximately 91% (n=212) of the participants reported being a US citizen or a permanent resident. About 35% (n=79) identified themselves as Chinese; the remainder were from other Asian countries.

As seen in Table 1, approximately half of the participants (n=122, 52%) identified themselves as male; 48% (n=112) were female. Further, 55% (n=128) of the participants reported being single, 42% (n=98) reported being married or being in a domestic partnership, and 3% (n=7) reported being widowed or divorced. About 6% (n=14) had an education level below college; 68% (n=160) reported having some college education, an associate degree, or a bachelor's degree; and 26% (n=60) had earned a master's or doctoral degree.

In total, 32% (n=72) of the participants were unemployed, 1% (n=3) was retired, and 67% (n=152) were employed on a part- or full-time basis at the time of data collection. Further, 23% (n=51) reported their household income as being <$39,999 in 2019, 49% (n=92) between $40,000 to $99,999, and 29% (n=64) >$100,000. In addition, 57% (n=126) of the participants reported that their family income had been impacted by the COVID-19 outbreak.

We also collected data on the participants' political beliefs and orientations: 18% (n=48) were Republicans, 47% (n=110) were Democrats, and 25% (n=58) were Independent. More than half of the participants (n=113, 52%) reported being liberal or very liberal, 27% (n=59) classified themselves as moderate, and 21% (n=46) were conservative or very conservative (Table 1).

Table 1. Participants' demographics (N=235).

Variable	Participants
Age (years)	
Mean (SD)	32.87 (10.95)
Range	18-73
Sex, n (%)	
Male	122 (52.1)
Female	112 (47.9)
Years in the United States	
Mean (SD)	23.56 (12.95)
Range	1-66
Education[a], n (%)	
Less than college	14 (5.6)
Some college or college degree	160 (68.4)
Graduate degree	60 (25.6)
Chinese, n (%)	
Yes	79 (34.8)
No	148 (65.2)
Citizen or permanent resident, n (%)	
Yes	212 (91)
No	21 (9)
Annual household income[a], n (%)	
≤$39,999	51 (22.8)
$40,000-$99,999	92 (48.7)
≥$100,000	64 (28.6)
Income impacted by COVID-19, n (%)	
Yes	126 (57.3)
No	94 (42.7)
Marital status, n (%)	
Single	128 (54.9)
Married or have a domestic partner	98 (42.1)
Widowed	2 (0.9)
Divorced	5 (2.1)
Employment, n (%)	
Part-time or full-time	152 (67)
Not employed	72 (31.7)
Retired	3 (1.3)
Party affiliation, n (%)	
Republican	48 (18.3)
Democrat	110 (46.8)
Independent	58 (24.7)
Political orientation[a], n (%)	
Liberal	113 (51.8)

XSL•FO
RenderX

Variable	Participants
Moderate	59 (27.1)
Conservative	46 (21.1)

[a]Regrouped for this table.

Prevalence of Experienced Discrimination

Our first question asked about the prevalence of various types of discrimination experienced by Asians in the United States (Table 2). Pooling the responses "very often" and "often," each experienced discrimination type ranged from 5% to 11%. The most frequently experienced discrimination types included "people act as if they think you are dangerous" (n=25, 11%); "being treated with less courtesy or respect" (n=24; 10%); followed by "received poorer service than other people" (n=20,

9%), and "people act if they are afraid of you" (n=17, 7%). Compared to the other four types of discriminatory experiences, encounters where individuals were threatened or harassed took place less frequently; only 5% (n=11) of participants experienced it "very often" or "often," 9% (n=21) reported "sometimes," and 86% (n=203) answered "rarely" or "never." When the responses "very often," "often," and "sometimes" were pooled, each discrimination type was experienced by 14%-34% of the sample.

Table 2. Experience of discrimination among survey participants (N=235).

Item	Very often, n (%)	Often, n (%)	Sometimes, n (%)	Rarely, n (%)	Never, n (%)
You are treated with less courtesy or respect than other people.	11 (4.7)	13 (5.5)	48 (20.4)	67 (28.5)	96 (40.9)
People act as if they think you are dangerous.	7 (3.0)	18 (7.7)	55 (23.4)	50 (21.3)	105 (44.7)
People act as if they are afraid of you.	6 (2.6)	11 (4.7)	60 (25.5)	52 (22.1)	106 (45.1)
You receive poorer service than other people	4 (1.7)	16 (6.8)	43 (18.3)	59 (25.1)	113 (48.1)
You are threatened or harassed.	3 (1.3)	8 (3.4)	21 (8.9)	51 (21.7)	152 (64.7)

Prevalence of Concerns About Future Discrimination

In terms of concerns about future experiences of discrimination, a higher percentage of the participants answered "very often," "often," or "sometimes" for the five discrimination types listed in Table 3. The results also showed that participants were generally concerned about the given forms of discrimination at similar levels. The pooled percentages of "very often" and "often" for each discrimination type ranged from 25% to 27%. The discrimination types of most frequent concern were "will be treated with less courtesy or respect than other people" (n=64, 27%) and "will receive poorer service than other people" (n=62, 26%). When the response "sometimes" was included, the range increased to 49%-58% (Table 3).

Table 3. Concerns about future discrimination among survey participants (N=235).

Item	Very often, n (%)	Often, n (%)	Sometimes, n (%)	Rarely, n (%)	Never, n (%)
You will be treated with less courtesy or respect than other people.	18 (7.7)	46 (19.6)	71 (30.2)	46 (19.6)	54 (23.0)
People will act as if they think you are dangerous.	20 (8.5)	39 (16.6)	67 (28.5)	45 (19.1)	64 (27.2)
People will act as if they are afraid of you	17 (7.2)	43 (18.3)	63 (26.8)	48 (20.4)	64 (27.2)
You will receive poorer service than other people	17 (7.2)	45 (19.1)	62 (26.4)	47 (20.0)	64 (27.2)
You will be threatened or harassed.	19 (8.1)	40 (17.0)	57 (24.3)	52 (22.1)	67 (28.5)

Factors Associated With Discrimination Experience and Concerns

Another question examined factors that may be associated with the experience of discrimination or concerns about future discrimination. The two sets of hierarchical regression analyses resulted in two significant models for experience of discrimination [R^2=.183, $F_{14,187}$=3.00, P<.001] and concerns about future discrimination [R^2=.166, $F_{14,187}$=2.66, P=.001].

Tables 4 and 5 demonstrate the results associated with these tests.

Our analyses revealed that age was negatively associated with both experience of discrimination (β=–.18, P=.03) and concerns about future discrimination (β=–.26, P=.003). This showed that younger people were more likely to experience discrimination and worry about future discrimination they may face. The length of residing in the United States was positively associated with both experience of discrimination (β=.20, P=.02) and concerns about future discrimination (β=.21, P=.02). This shows that the

longer people lived in the United States, the more likely they felt discriminated or worried about how they might be treated in the future. Among different Asian ethnic groups residing in the United States, Chinese people were more likely to experience discrimination (β=.18, P=.01) or have concerns about future discrimination than other Asian groups (β=.15, P=.04).

Individuals who often used print (β=.31, P<.001) or social media (β=.18, P=.01) to obtain COVID-19–related news reported having experienced more discrimination. Using social media for COVID-19 was also positively associated with greater concerns about future discrimination (β=.20, P=.005).

Table 4. Hierarchical regression analysis for variables relating to experience of discrimination. Italics indicates significant results.

Step and variable	Model 1		Model 2		R^2 change	F change	P value
	β[a]	P value	β[a]	P value			
Step 1: Demographic variables					.077	1.77	.08
Age	–.13	.14	*–.18*	*.03*			
Years in the United States	*.18*	*.045*	*.20*	*.02*			
Sex	–.00	.96	–.02	.81			
Chinese	*.19*	*.01*	*.18*	*.01*			
Visitor	–.01	.91	–.04	.54			
Education	.04	.59	.06	.42			
Employment	–.04	.59	–.03	.64			
Income	–.08	.31	–.05	.47			
Political orientation	.05	.52	.09	.26			
Step 2: COVID-19 media source					.107	4.88[b]	.001
Print	__[c]	—	*.31*	*<.001*			
Radio	—	—	.02	.79			
TV	—	—	–.07	.37			
News website or app	—	—	–.05	.48			
Social media	—	—	*.18*	*.013*			

[a]All beta coefficients are standardized regression coefficients.

[b]F change is significant at P<.001.

[c]Not applicable.

Table 5. Hierarchical regression analysis for variables relating to concerns about future discrimination. Italics indicates significant results.

Step and variable	Model 1		Model 2		R^2 change	F change	P value
	β[a]	P value	β	P value			
Step 1: Demographic variables					.114	2.74	.005
Age	–.27	.002	–.26	.003			
Years in the United States	.17	.047	.21	.02			
Sex	–.03	.73	–.04	.56			
Chinese	.15	.04	.15	.04			
Visitor	–.03	.68	–.04	.55			
Education	–.01	.89	–.02	.83			
Employment	–.08	.29	–.07	.37			
Income	.03	.71	.03	.71			
Political orientation	–.11	.11	–.10	.17			
Step 2: COVID-19 media source					.052	2.34	.04
Print	—[b]	—	.07	.39			
Radio	—	—	-.01	.92			
TV	—	—	.04	.59			
News website or app	—	—	.06	.38			
Social media	—	—	.20	.005			

[a]All beta coefficients are standardized regression coefficients.

[b]Not applicable.

Discussion

Principal Findings

This study examined what types of discrimination Asians residing in the United States have experienced and worried about during the pandemic, as well as factors that were associated with everyday discrimination experience and concerns about future discrimination that the Asian community may face. Our study suggested that a substantial part of Asians surveyed had experienced and worried about some type of discrimination. The relationship between media sources and the perception of racial biases among this group was also identified. Our study discovered the significant role of social media use in the higher likelihood of experiencing discrimination and having concerns about future discrimination. Use of print media was also positively associated with experiencing discrimination. The detailed discussion of these findings is presented below.

First, our study provided additional empirical evidence suggesting that a substantial part of the Asian community in the United States had experienced and worried about some type of discrimination at least part of the time (as indicated by the response "sometimes"). These findings complemented what the Center for Public Integrity Poll had suggested: 1 out of 4 people in the United States said they would be concerned about close contact with someone of Asian ancestry in public during the pandemic [5,12]. Additionally, about 14% (32/235) of the participants reported very often, often, or sometimes being threatened or harassed. These numbers are very concerning given that Asian Americans account for 6% of the total

population in the United States, representing the fastest growing ethnic group [31]. These numbers are not negligible because disease-related anti-Asian sentiment has caused some damage—for example, an Asian American family was stabbed in March in Texas while shopping at a Sam's Club; an Asian woman was hit on the head with an umbrella in mid-April accompanied with anti-Asian insults [5]. These types of hate crimes have caused additional fear and anxiety among Asians [7]. Our study further suggests that during a prolonged pandemic like COVID-19, we should pay careful attention to other social issues relating to the pandemic.

Second, frequent use of social media to obtain COVID-19 news was positively associated with both experience of everyday discrimination and concerns about future discrimination. This finding can be explained by a number of reasons. First, social media can be easily flooded with unconfirmed negative stereotypes, emotion-arousing information, and racist hate speech [17,32]. Therefore, those who use social media more frequently may have a higher chance of being exposed to racism-related information. Second, research has shown that social media is usually associated with a high degree of ethnic homophily, meaning that people from the same ethnic group are more likely to gather together on social media [33,34]. Our Asian participants likely had a social media network composed of a relatively high percentage of Asian friends, who may have shared much racism-related information as it had become a primary concern in the Asian community. Finally, social media algorithms present information to users based on users' earlier behaviors [35]. It is conceivable that, through social media

XSL•FO
RenderX

algorithms, those who frequently resort to social media for COVID-19 news are then exposed to more COVID-19 news, which may include reports on anti-Asian incidents (see examples in [4,5,12]). Thus, Asian participants who frequently accessed COVID-19 news on social media may have been overinformed of and hypervigilant about COVID-19–related racism, and thus reported more experiences of and concerns about discrimination. One recent study showed that Asians/Asian Americans who engaged in social media more during the pandemic reported more worry about being discriminated, likely because social media browsing during this time led to the perception that anti-Asian racism was pervasive [18], which resonates with our findings.

Third, it is noteworthy that usage of print media was also positively associated with experience of discrimination. Research has found some patterns regarding how different media might affect people's information processing [36]. For example, it is noted that the nature of print media in dominantly delivering text-based information and allowing readers to process information at their own pace and sequence increases the chances of elaborating information in a deeper cognitive level, especially for those who have higher interest in the information [37]. On the other hand, television is more powerful in involving those who are less involved in the content because of its ability of activating multiple sensory modes [38]. Being treated as the main target of discrimination during the COVID-19 pandemic, Asians would be highly involved in racism-related news coverage. In this case, print media could facilitate more cognitive elaboration of relevant information among Asians, which might lead them to be more sensitive to the discrimination they experienced in their daily lives.

Finally, our study also indicated that those who self-identified as Chinese residing in the United States were more likely to experience discrimination and have concerns about future discrimination. It is not surprising that Chinese individuals were more sensitive to discrimination considering the close connection between the COVID-19 outbreak and China. Furthermore, we also found that younger age and a lengthier residency in the United States were both associated with experience of and concerns about discrimination. It would be worthwhile to conduct qualitative research to find out how age and the length of time spent living in a different country (and cultural context) can cause different perceptions of racial prejudice and discrimination.

Our study supports earlier reports that demonstrate the concerning state of Asian people during the pandemic. It reveals the potential psychological toll of attributing a pandemic to a particular ethnic group. Our results can be useful for public health agencies. To reduce discrimination against Asians during the pandemic, official sources and public health professionals should be cognizant of the possible impacts of amplifying stigmatizing cues in media reports on activating racial biases. Furthermore, health communication efforts can further emphasize evidence-based prevention measures of curbing the pandemic and help remedy social prejudices relating to the disease. Finally, Asian users could also be informed that using social media for COVID-19 information would associate with greater concerns about future discrimination, and thus they may consider approaching this media source with caution.

Limitations and Future Studies

As an exploratory examination of Asians and their experienced and expected discrimination during COVID-19, the study has some limitations. First, even though this project reached Asians from 35 states in the country, the nonprobability sampling nature of this study will limit generalizations of the results to the whole Asian community. Second, our study revealed the prevalence of discrimination against Asians during COVID-19, but the nature of the study did not allow us to report whether and by how much discrimination increased since the outbreak. Scholars who have collected longitudinal data before and after COVID-19 are in a better position to present the change. Third, we used an existing scale of racial discrimination [24], which generated self-report data. Future studies can explore new ways to better assess the presence of racial discrimination, using experimental methods or attitudinal and behavioral indicators, as suggested by the National Research Council [39]. Lastly, our sample included Asians who resided in the United States during the outbreak and comprised US citizens, permanent residents, and temporary visitors. We asked our participants whether they self-identified as Chinese but did not ask the non-Chinese participants to report their original nationality. Future scholars may collect this information to analyze the potential role of nationality in influencing individual experiences during a global pandemic.

Conclusions

Through a cross-sectional survey of Asian people in the United States, our study showed that a substantial percentage of this group has experienced and worried about some form of discrimination. Use of print and social media for COVID-19 news were both associated with a higher likelihood of experiencing discrimination, and social media use was further associated with concerns about future discrimination. Notably, our study contributes to the understanding of the media's role in shaping racism-related perceptions of a group experiencing discrimination during a global health crisis. These findings revealed the importance of addressing discrimination in a health pandemic to protect the well-being of minority groups that are being linked to infectious diseases.

XSL•FO

RenderX

Conflicts of Interest

None declared.

References

1. Hernandez S, O'Key S, Watts K, Manley B, Pettersson H. Tracking COVID-19 cases in the US. CNN. 2020. URL: https://www.cnn.com/interactive/2020/health/coronavirus-us-maps-and-cases/ [accessed 2020-09-23]

2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-09-23]

3. Mayor E, Eicher V, Bangerter A, Gilles I, Clémence A, Green EGT. Dynamic social representations of the 2009 H1N1 pandemic: Shifting patterns of sense-making and blame. Public Underst Sci 2013 Nov;22(8):1011-1024. [doi: 10.1177/0963662512443326] [Medline: 23825240]

4. Mak G. Being Asian-American in the time of COVID-19. The Guardian. 2020 May 8. URL: https://www.theguardian.com/commentisfree/2020/may/08/asian-american-coronavirus-geoffrey-mak [accessed 2020-06-10]

5. Campbell AF, Ellerbeck A. Federal agencies are doing little about the rise in anti-Asian hate. The Center for Public Integrity. 2020 Apr 16. URL: https://publicintegrity.org/health/coronavirus-and-inequality/federal-agencies-are-doing-little-about-the-rise-in-anti-asian-hate/ [accessed 2020-08-11]

6. In one month, STOP AAPI Hate received almost 1500 incident reports of verbal harassment, shunning and physical assaults. Asian Pacific Policy & Planning Council. 2020 Apr 24. URL: http://www.asianpacificpolicyandplanningcouncil.org/wp-content/uploads/Press_Release_4_23_20.pdf [accessed 2020-08-11]

7. Gover AR, Harper SB, Langton L. Anti-Asian hate crime during the COVID-19 pandemic: Exploring the reproduction of inequality. Am J Crim Justice 2020 Jul 07:1-21 [FREE Full text] [doi: 10.1007/s12103-020-09545-1] [Medline: 32837171]

8. Mallin A, Margolin J. Homeland security warns terrorists may exploit COVID-19 pandemic. ABC News. 2020 Mar 24. URL: https://abcnews.go.com/Politics/homeland-security-warns-terrorists-exploit-covid-19-pandemic/story?id=69770582 [accessed 2020-04-07]

9. Dhanani L, Franz B. Unexpected public health consequences of the COVID-19 pandemic: A national survey examining anti-Asian attitudes in the USA. Int J Public Health 2020 Jul;65(6):747-754 [FREE Full text] [doi: 10.1007/s00038-020-01440-0] [Medline: 32728852]

10. Misra S, Le PD, Goldmann E, Yang LH. Psychological impact of anti-Asian stigma due to the COVID-19 pandemic: A call for research, practice, and policy responses. Psychol Trauma 2020 Jul;12(5):461-464. [doi: 10.1037/tra0000821] [Medline: 32525390]

11. Covid-19 fueling anti-Asian racism and xenophobia worldwide. Human Rights Watch. 2020 May 12. URL: https://www.hrw.org/news/2020/05/12/covid-19-fueling-anti-asian-racism-and-xenophobia-worldwide [accessed 2020-06-10]

12. Ellerbeck A. Over 30 percent of Americans have witnessed COVID-19 bias against Asians, poll says. NBC News. 2020 Apr 28. URL: https://www.nbcnews.com/news/asian-america/over-30-americans-have-witnessed-covid-19-bias-against-asians-n1193901 [accessed 2020-06-10]

13. Croucher S, Nguyen T, Rahmani D. Prejudice toward Asian Americans in the Covid-19 pandemic: The effects of social media use in the United States. Front. Commun 2020 Jun 12;5:39 [FREE Full text] [doi: 10.3389/fcomm.2020.00039]

14. Mental Health Fund. Crisis Text Line. 2020. URL: https://www.crisistextline.org/ [accessed 2020-06-10]

15. Reny TT, Barreto MA. Xenophobia in the time of pandemic: Othering, anti-Asian attitudes, and COVID-19. Politics, Groups, and Identities 2020 May 28:1-24. [doi: 10.1080/21565503.2020.1769693]

16. Davidson T, Farquhar L. Prejudice and social media: Attitudes toward illegal immigrants, refugees, and transgender people. In: Farris D, Compton D, Herrera A, editors. Gender, Sexuality and Race in the Digital Age. Cham: Springer; 2020:151-167.

17. Ziems C, He B, Soni S, Kumar S. Racism is a virus: Anti-Asian hate and counterhate in social media during the COVID-19 crisis. arXiv. Preprint posted online May 25, 2020 2020 [FREE Full text]

18. Yang C, Tsai J, Pan S. Discrimination and well-being among Asians/Asian Americans during COVID-19: The role of social media. Cyberpsychol Behav Soc Netw 2020 Aug 07 [FREE Full text] [doi: 10.1089/cyber.2020.0394] [Medline: 32762541]

19. Gerbner G, Gross L. Living with television: the violence profile. J Commun 1976;26(2):173-199. [doi: 10.1111/j.1460-2466.1976.tb01397.x] [Medline: 932235]

20. Oliver M, Ramasubramanian S, Kim J. Media and racism. In: Roskos-Ewoldsen DR, Monahan JL, editors. Communication and Social Cognition: Theories and Methods. New York: Routledge; Mar 2007:273-294.

21. Bissell K, Parrott S. Prejudice: the role of the media in the development of social bias. J Commun Monogr 2013 Nov 18;15(4):219-270. [doi: 10.1177/1522637913504401]

22. Nairn R, Pega F, McCreanor T, Rankine J, Barnes A. Media, racism and public health psychology. J Health Psychol 2006 Mar;11(2):183-196. [doi: 10.1177/1359105306061179] [Medline: 16464918]

23. Williams DR, Mohammed SA. Discrimination and racial disparities in health: Evidence and needed research. J Behav Med 2009 Feb;32(1):20-47 [FREE Full text] [doi: 10.1007/s10865-008-9185-0] [Medline: 19030981]

24. Williams DR, Yan Yu, Jackson JS, Anderson NB. Racial differences in physical and mental health: Socio-economic status, stress and discrimination. J Health Psychol 1997 Jul;2(3):335-351. [doi: 10.1177/135910539700200305] [Medline: 22013026]

25. Williams DR, Wyatt R. Racial bias in health care and health: Challenges and opportunities. JAMA 2015 Aug 11;314(6):555-556. [doi: 10.1001/jama.2015.9260] [Medline: 26262792]

26. Harnois CE, Bastos JL, Campbell ME, Keith VM. Measuring perceived mistreatment across diverse social groups: An evaluation of the Everyday Discrimination Scale. Soc Sci Med 2019 Jul;232:298-306. [doi: 10.1016/j.socscimed.2019.05.011] [Medline: 31121440]

27. Hauser D, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav Res Methods 2016 Mar;48(1):400-407. [doi: 10.3758/s13428-015-0578-z] [Medline: 25761395]

28. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: Cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]

29. Park S, Boatwright B, Johnson Avery E. Information channel preference in health crisis: Exploring the roles of perceived risk, preparedness, knowledge, and intent to follow directives. Public Relations Review 2019 Dec;45(5):101794. [doi: 10.1016/j.pubrev.2019.05.015]

30. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 2009 Nov;41(4):1149-1160 [FREE Full text] [doi: 10.3758/brm.41.4.1149]

31. López G, Ruiz NG, Patten E. Key facts about Asian Americans, a diverse and growing population. Pew Research Center. 2017 Sep 8. URL: https://www.pewresearch.org/fact-tank/2017/09/08/key-facts-about-asian-americans/ [accessed 2020-08-17]

32. Bontcheva K, Gorrell G, Wessels B. Social media and information overload: Survey results. arXiv. Preprint posted online Jun 4, 2013 2020 [FREE Full text]

33. Mazur E, Richards L. Adolescents' and emerging adults' social networking online: Homophily or diversity? J Appl Dev Psychol 2011 Jul;32(4):180-188. [doi: 10.1016/j.appdev.2011.03.001]

34. Thelwall M. Homophily in MySpace. J Am Soc Inf Sci 2009 Feb;60(2):219-231. [doi: 10.1002/asi.20978]

35. Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science 2015 Jun 05;348(6239):1130-1132. [doi: 10.1126/science.aaa1160] [Medline: 25953820]

36. Dijkstra M, Buijtels HE, van Raaij W. Separate and joint effects of medium type on consumer responses: a comparison of television, print, and the Internet. Journal of Business Research 2005 Mar;58(3):377-386. [doi: 10.1016/s0148-2963(03)00105-x]

37. Jacoby J, Hoyer W, Zimmer M. To read, view, or listen? A cross-media comparison of comprehension. Current Issues and Research in Advertising 1983;6(1):201-217 [FREE Full text] [doi: 10.1080/01633392.1983.10505339]

38. Buchholz LM, Smith RE. The role of consumer involvement in determining cognitive response to broadcast advertising. Journal of Advertising 1991 Mar;20(1):4-17. [doi: 10.1080/00913367.1991.10673202]

39. National Research Council. Measuring Racial Discrimination. Washington, DC: National Academies Press; 2004.

Abbreviations

IRB: institutional review board
MTurk: Mechanical Turk
SARS: severe acute respiratory syndrome

Original Paper

Perception of the COVID-19 Pandemic Among Patients With Inflammatory Bowel Disease in the Time of Telemedicine: Cross-Sectional Questionnaire Study

Fabiana Zingone[1], MD, PhD; Monica Siniscalchi[2], LicPs; Edoardo Vincenzo Savarino[1], MD, PhD; Brigida Barberio[1], MD; Linda Cingolani[1], MD; Renata D'Incà[1], MD; Francesca Romana De Filippo[2], MD; Silvia Camera[2], MD; Carolina Ciacci[2], MD

[1]Inflammatory Bowel Disease Center, Department of Surgery, Oncology, and Gastroenterology, University of Padua, Padua, Italy

[2]Inflammatory Bowel Disease Center, Department of Medicine, Surgery, Dentistry, Scuola Medica Salernitana, University of Salerno, University Hospital San Giovanni di Dio e Ruggi d'Aragona, Salerno, Italy

Corresponding Author:
Carolina Ciacci, MD
Inflammatory Bowel Disease Center
Department of Medicine, Surgery, Dentistry, Scuola Medica Salernitana, University of Salerno
University Hospital San Giovanni di Dio e Ruggi d'Aragona
Via San Leonardo 1
Salerno, 84131
Italy
Phone: 39 089672635
Email: cciacci@unisa.it

Abstract

Background: After the COVID-19 outbreak, the Italian Government stopped most regular health care activity. As a result, patients with inflammatory bowel disease (IBD) had limited access to outpatient clinics and hospitals.

Objective: This study aimed to analyze the perception of the COVID-19 emergency among patients with IBD during the early weeks of the lockdown.

Methods: We invited adult patients with IBD from the University of Salerno (Campania, South Italy) and the University of Padua (Veneto, North Italy) by email to answer an ad hoc anonymous survey about COVID-19. We also collected data on demographic and disease characteristics.

Results: In total, 167 patients with IBD from Padua and 83 patients from Salerno answered the survey (age: mean 39.7 years, SD 13.9 years; female: n=116, 46.4%). We found that patients with IBD were particularly worried about the COVID-19 pandemic (enough: 77/250, 30.8%; much/very much: 140/250, 56.0%), as they felt more vulnerable to COVID-19 due to their condition (enough: 70/250, 28.0%; much/very much: 109/250, 43.6%). Patients with IBD from the red zone of Veneto were more worried than patients from Campania (P=.001), and men felt more susceptible to the virus than women (P=.05). Additionally, remote medicine was appreciated more by younger patients than older patients (P=.04).

Conclusions: The results of our survey demonstrate that the lockdown had a significant impact on the psychological aspects of patients with IBD and suggest the need for increasing communication with patients with IBD (eg, through telemedicine) to ensure patients receive adequate health care, correct information, and proper psychological support.

Introduction

The COVID-19 pandemic is a public health emergency [1]. Studies from China, the country where the SARS-CoV-2 outbreak initially occurred, have indicated that the potential risk factors for poor prognoses in people who contract the SARS-CoV-2 virus include male gender, older age, high Sequential Organ Failure Assessment scores, and high D-dimer

levels [2,3]. The extraordinary measures that have been implemented to contain viral spread, such as the cancellation of flights, the initial lockdown of large areas in China, and the subsequent lockdown of large areas in Korea and Italy, have captured the attention and interest of the public. However, these measures have also generated misconceptions and fear [4].

In Italy, the first person-to-person virus transmission was reported by authorities on February 21, 2020. One of the first places to experiment with quarantine and lockdown was Vo', a small town near Padua in Veneto, on February 23, 2020. Subsequently, after a few days, lockdown measures were extended to the whole country. Outside of the lockdown measures adopted in China to combat the outbreak, the lockdown measures adopted in Italy were considered to be the most radical. Italians were not allowed to leave their houses. This excluded those who worked in essential services, such as health care, transportation, and supermarkets. Additionally, only one person per family was allowed to go out for food shopping. Further, while outside, the Italian population was asked to wear surgical masks and gloves. Some regions of Northern Italy, such as Veneto, were defined as red zones due to the high number of people infected with SARS-CoV-2 and deaths caused by SARS-CoV-2 infection. Moreover, the mass media and the internet have put significant pressure on the population due to the potentially lethal implications of SARS-CoV-2 infection. Therefore, due to the restrictions that have been implemented since the beginning of the quarantine, public and private health care workers have only been able to address emergencies.

Patients with inflammatory bowel disease (IBD) are at a higher risk of infection than the general population. This is mainly due to the immunosuppressive therapy that patients with IBD receive, which involves steroids, thiopurines, and biologic agents [5-8]. A large study demonstrated that patients with IBD had an increased risk of influenza infection and were more likely to require hospitalization than those without IBD [5]. However, the magnitude of this increased risk of infection and how it relates to the use of available medical therapies remains controversial [9]. Some studies have indicated the possibility of major life events inducing relapses in patients with IBD [10,11]. During the COVID-19 pandemic, patients with IBD have had to limit their access to outpatient clinics and hospitals. However, a vast majority of patients with IBD at our outpatient clinics have continued their immunosuppressant or biologic therapy. Health care providers have had to rapidly and abruptly convert their traditional methods for practicing medicine to the more modern concept of telemedicine [12]. For instance, phone calls or internet video calls via patient-available apps (eg, Skype, WhatsApp, Zoom) and emails have been used to contact patients with IBD and continue their follow-up. The aim of this study

is to report the perception of the COVID-19 emergency among a cohort of patients with IBD who live in a red zone during the early weeks of lockdown and their ability to access remote consultations in order to improve their health care during possible future emergencies similar to the COVID-19 pandemic.

Methods

During quarantine, personnel from IBD centers advised patients to postpone their regular outpatient visits if patients' reasons for visiting were not urgent. They also invited patients to have remote health care consultations. With regard to regular pre-COVID-19 face-to-face outpatient visits in Italy, patients used to receive a report regarding drug prescriptions and lab tests, which patients were to return during the next follow-up visit. With telemedicine, patients in Italy are given their lab test results by email or fax prior to remote health care consultation. These results, along with patient evaluations for general health status, are then discussed during follow-up visits performed via remote consultation

Between March 14 and 24, 2020, we invited patients with IBD from the IBD Units of the University of Salerno (Campania region, South Italy) and the University of Padua (Veneto region, North Italy) (Figure 1) by email to answer an ad hoc anonymous survey about COVID-19. Similar surveys have been used at our centers for two other chronic conditions [13,14]. Patients provided their email address and consented to contact. Therefore, the survey used in this study follows the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist [15]. Patients enrolled from both universities only included adult patients with a confirmed IBD diagnosis and an equal distribution of men and women. The anonymous web survey included 14 multiple-choice questions that aimed to evaluate patients' perception of the COVID-19 pandemic. In particular, we asked if patients were worried about the epidemic. The survey included questions about whether patients felt more susceptible to SARS-CoV-2 infection compared to the general population, whether they preferred to avoid going to the hospital, and their thoughts about remote consultation. We recorded patients' sex, date of birth, time of IBD diagnosis, region of origin (Veneto or Campania), and type of disease (ie, Crohn disease, ulcerative colitis, and unclassified disease) [16]. We also asked for information regarding patients' ongoing therapy and comorbidities. Patients also reported whether they had received the influenza vaccination. We expected an average time of 10 minutes to complete the survey. We sent a gentle reminder 5 days after the request was sent. We closed the survey to patients 10 days after it was sent out and analyzed the survey data.

Figure 1. Map of Italy detailing the spread of COVID-19. The map was last updated on March 24, 2020 for Veneto and Campania.
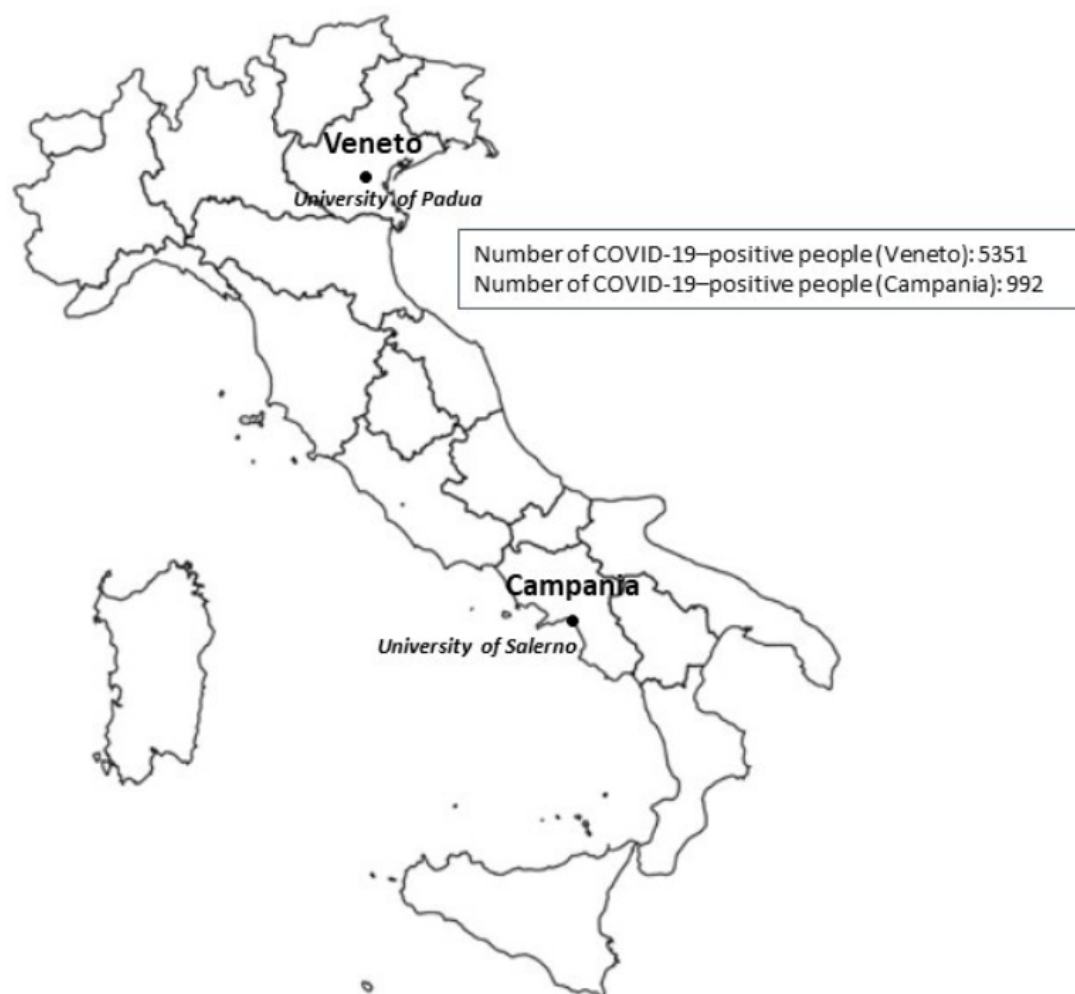
Categorical and continuous variables were expressed as proportions with percentages and means with standard deviations, respectively. Comparisons between categorical variables were performed using the Chi-square test. A *P* value of <.05 was considered statistically significant. STATA 11 software was used for statistical analysis.

Results

Of the 450 surveys sent, 250 (55.5%) were answered. In total, 167 patients from Veneto (age: mean 39.1 years, SD 13.5 years; male: n=91, 54.5%) and 83 patients from Campania (age: mean 39.1 years, SD 14.8 years; male: n=43, 51.8%) answered the survey. There were no statistically significant differences in age and sex between patients from Veneto and those from Campania. Of the 250 patients who answered the survey, 142 (56.8%) had Crohn disease and 100 (40.0%) had ulcerative colitis. The mean time from diagnosis to the point of the survey was 10.7 years (SD 10.6 years). More than half of our study population was on immunosuppressive/immunomodulatory therapy (138/250, 55.2%). Less than 15% (36/250) of patients had at least one other chronic disease, including diabetes, high blood pressure, and dyslipidemia (Table 1).

Table 1. Study population characteristics (N=250).

Characteristic	Participants
Region of origin, n (%)	
Campania	83 (33.2)
Veneto	167 (66.2)
Sex – male, n (%)	134 (53.6)
Age (years), mean (SD)	39.7 (13.9)
Age groups (years), n (%)	
<35	104 (41.6)
35-49	82 (32.8)
≥50	64 (25.6)
Time from diagnosis (years), mean (SD)	10.7 (10.6)
Time from diagnosis by age group (years), mean (SD)	
<5	87 (34.8)
5-9	46 (18.4)
≥10	117 (46.8)
Diseases, n (%)	
Ulcerative colitis	100 (40)
Crohn disease	142 (56.8)
Unclassified disease	8 (3.2)
Ongoing Therapy, n (%)	
Mesalazine or no therapy	112 (44.8)
Steroids, azathioprine, or biologics	138 (55.2)
Chronic disease[a] , n (%)	
No	214 (85.6)
Yes	36 (14.4)

[a]At least one other chronic disease, including diabetes, high blood pressure, and dyslipidemia.

Table 2 and 3 shows the web survey answers of the study population (N=250). Most patients were worried about COVID-19 (for Question 1, enough: n=77, 30.8%; much/very much: n=140, 56.0%) because of social distancing or going to crowded places, such as supermarkets and food shops (for Question 3, enough: n=32, 17.2%; much/very much: n=164, 65.6%). Many patients also felt disturbed and tense when thinking of SARS-CoV-2 infection (for Question 5, enough: n=77, 30.8%; much/very much: n=115, 46.0%). Moreover, patients with IBD felt more vulnerable to COVID-19 (for Question 2, enough: n=70, 28.0%; much/very much: n=109, 43.6%) and depressed (for Question 7, much/very much: n=22, 88.8%) due to their condition. Most patients (148/250, 59.2%) did not want to go to outpatient clinics at the hospital, and more than 70% (187/250) of patients preferred to undergo remote medical examination.

We tested for differences in survey answers between subgroups. Based on patients' answers to Question 7 ("I feel depressed because of my disease"), younger patients (<35 years) felt significantly more depressed (answered "much" or "very much") than older patients (≥50 years) (age: <35 years: 96/104, 92.3%;

35-50 years: 73/82, 89.02%; ≥50 years: 53/64, 82.82%; P=.04). Moreover, based on patients' answers to Question 9 ("Are you happy with telemedicine remote visits?"), younger patients preferred remote medical control (answered "yes, perfect") significantly more than the older patients (age: <35 years: 80/104, 76.9%; 35-50 years: 64/82, 78.05%; ≥50 years: 43/64, 67.2%; P=.04). Figure 2 shows that patients from the red zone of Veneto were more worried about SARS-CoV-2 infection (Question 1: P=.001), felt that they were at a higher risk of infection due to IBD (Question 2: P=.003), were more afraid of going to crowded places (Question 3: P=.001), were more tense and disturbed thinking about COVID-19 (Question 5: P=.001), and were more depressed due to their condition (Question 7: P=.007) compared to patients from Campania. Women were more worried (Question 1: P=.049) and felt more disturbed and tense thinking about COVID-19 (Question 5: P=.003) than men. However, men felt more susceptible to the virus due to their therapy (Question 10: P=.05) (Figure 3). More immunosuppressed patients thought that infection might interfere with their therapy (for Question 6, much/very much: 73/138, 52.9% vs 36/112, 32.1% P=.002) and felt more susceptible to the infection due to their therapy (for Question

10, yes: 77/138, 55.8% vs 44/112, 39.3%), $P=.019$) than patients who were not immunosuppressed. We did not find any significant differences in survey answers regarding the presence of other chronic diseases and the time from diagnosis to the point of the survey. Only one-third (83/250) of the study population received the influenza vaccination in the previous 5 months, and 47.2% (118/250) of patients reported that they would receive a vaccination against COVID-19 if available. However, a similar percentage (117/250, 46.8%) of patients had doubts about COVID-19 vaccination.

Table 2. Results of the survey administered to patients with inflammatory bowel disease for evaluating COVID-19 perception (Questions 1-9).

Question	Not at all (%)	A little (%)	Enough (%)	Much (%)	Very much (%)
1. How much are you worried because of the coronavirus 19 pandemic?	2	11.2	30.8	30	26
2. Do you think that you are at higher risk of coronavirus 19 infection because you have IBD?	10	18.4	28	25.2	18.4
3. Are you worried because of social distancing or of going to crowded places, such as supermarkets, food shops?	1.6	15.6	17.2	29.6	36
4. Do you think that the coronavirus 19 information is excessive?	40.4	23.6	20.8	9.6	5.6
5. Do you feel disturbed or tense thinking about coronavirus infection?	5.2	18	30.8	26.4	19.6
6. Are you worried that the coronavirus infection may interfere with your own therapy?	17.2	15.6	23.6	19.6	24
7. I feel depressed because of my disease	0.8	3.6	6.8	34	54.8
8. I am threatened by having the disease	1.6	5.6	16	47.6	29.2
9. Are you happy with telemedicine remote visits?	14 (I would like to speak with the doctors)	5.2 (No, I do not feel like being in care)	74.8 (Yes, perfect)	6 (I'm afraid I can't answer everything)	0

Table 3. Results of the survey administered to patients with inflammatory bowel disease for evaluating COVID-19 perception (Questions 10-14).

Question	No (%)	Yes (%)	Maybe (%)
10. Do you think that for your therapy you are more susceptible to the coronavirus infection compared to the general population?	16.8	48.4	34.8
11. Are you afraid that pandemic reduces your care by physicians to less than it would be necessary?	48	29.2	22.8
12. Are you reluctant to go to hospital, because of coronavirus infection?	28.8	59.2	12
13. Did you undergo seasonal flu vaccination?	67.6	32.4	0
14. Would you like to undergo a vaccination for coronavirus, when it becomes available?	6	47.2	46.8

Figure 2. Comparison of COVID-19 survey results between participants from different regions of origin (Veneto vs Campania) for Questions 1-3 (A-C).
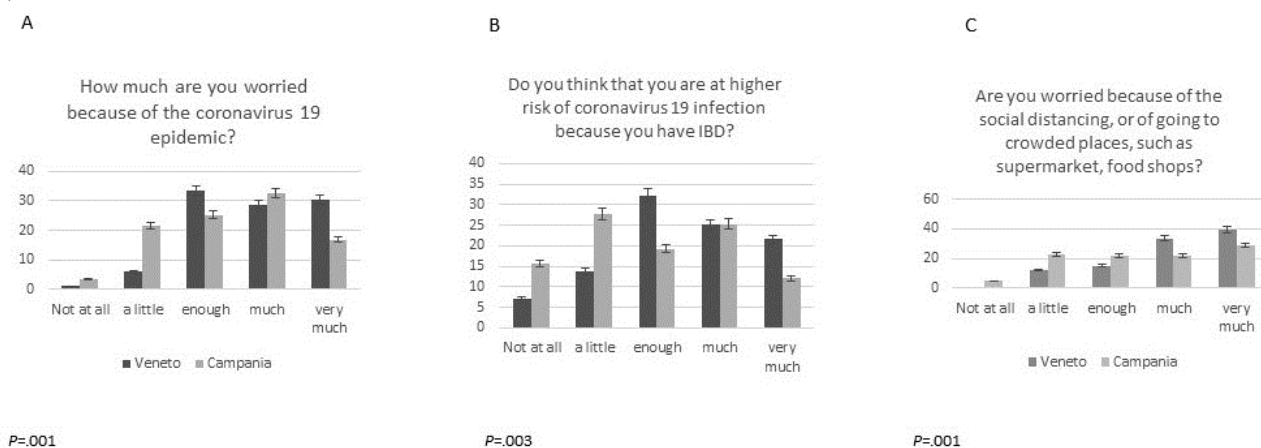
Figure 3. Comparison of COVID-19 survey results between male and female participants for Questions 1, 5, and 10 (A-C).
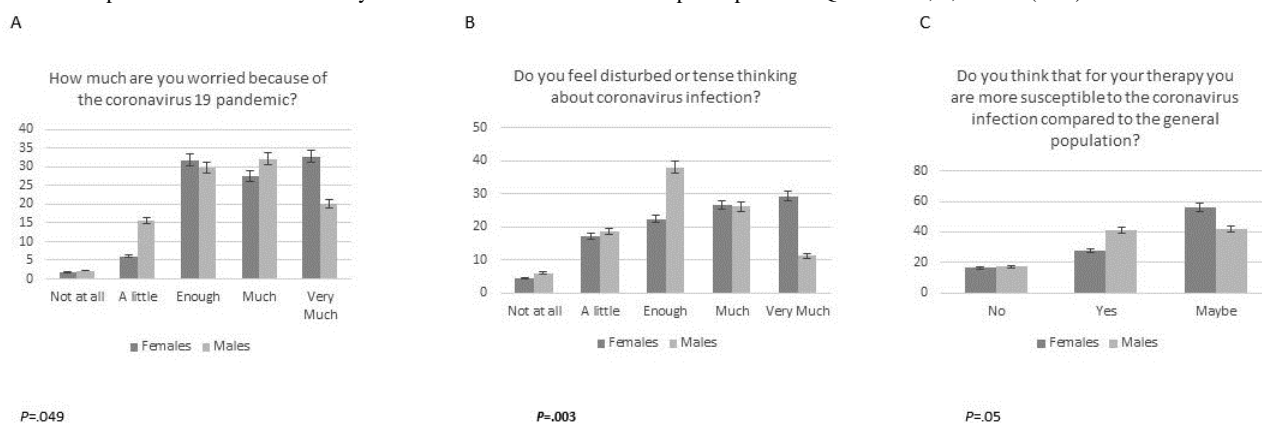
Discussion

This cross-sectional study reports on patients with IBD from two Italian regions and their perception of the COVID-19 pandemic during the quarantine. We found that patients with IBD were particularly worried, as they felt more vulnerable to SARS-CoV-2 infection due to their chronic disease and their immunosuppressant therapy. Patients with IBD have reported a low quality of life, disability, and poor sleep quality [17-19]. Thus, the COVID-19 pandemic has worsened an already precarious condition. Further, the unfiltered amount of information from social media during the days of the epidemic has confused and scared the Italian population [20].

Patients with IBD are likely to feel that they are at a higher risk of infection compared to the general population [5]. Our data indicated that patients with IBD from the Veneto region were more worried about SARS-CoV-2 infection than patients from Campania. Veneto was one of the first Italian regions where SARS-CoV-2 infection was initially observed, and it remains one of the regions with a high number of COVID-19–related deaths (Figure 1) [21]. Therefore, the high local spread of COVID-19 has likely affected the perception of patients.

Our survey showed similar results to those of another study, which indicated that a 30-day isolation for trained personnel, such as astronauts, do not have a significant impact on brain activity, neurotrophic factors, cognition, and mood, even though the stress levels of trained personnel significantly increased during isolation [22]. However, it is conceivable that behavioral and organic consequences may follow social isolation due to COVID-19 social distancing measures [23,24]. For example, similarly to other coronaviruses, changes in endocrine and immune response may occur, as millions of people worldwide are isolated in quarantine for minimizing the transmission of SARS-CoV-2, and social isolation can lead to neuroendocrine-immune changes [25].

According to the European Crohn's and Colitis Organisation, patients with IBD should have followed the national recommendations. However, some extra caution was needed. The recommendations published during the beginning of the pandemic suggested postponing the start of treatment with immunosuppressive drugs and biologics whenever possible [26,27]. When this was not possible, screening for SARS-CoV-2 infection should have been performed before starting biologics [28,29]. Moreover, to reduce access to hospitals and transport, the postponement of any elective or routine follow-up was recommended, along with the alternative to perform remote medical examinations [12,30]. Our study showed that most patients would be happy with remote medical examinations. However, older patients were not as happy with this alternative.

Our survey showed that men felt more susceptible to SARS-CoV-2 infection than women, but women felt more worried and tense about infection. This is likely due to both the country's National Health Institute and the media simultaneously reporting that men represent nearly 60% of people who tested positive for the SARS-CoV-2 virus and more than 70% of those who were infected with SARS-CoV-2 have died.

About two-thirds of our sample reported they had not received the influenza vaccination. This is relevant because the current guidelines recommend that patients with IBD should receive an influenza vaccination every year and pneumococcal vaccination with a booster every 5 years [31]. However, our percentage of patients who were immunized for influenza was similar to those in other studies [32]. Some patients also mentioned their fear of flu vaccination and vaccinations in general. This is likely in line with the high percentage of subjects that answered "maybe" to Question 14, which asks about receiving a future SARS-CoV-2 vaccination.

Our survey also shows that patients with IBD perceived immunosuppressive and immunomodulatory therapy as an added risk to general infection. Some studies have supported this concept, reporting a higher risk of severe and opportunistic infections in patients with IBD compared to control patients [6,7]. However, other data have not suggested an increased risk in infection [9]. Patients with IBD often ask their doctors about their risk of infection, how they should behave, and what measures they should take. A multicenter study conducted in Italy showed that active IBD, old age, and comorbidities were associated with poor COVID-19 outcomes, whereas IBD treatments were not [33].

Our study had several strengths. First, to the best of knowledge, our study is the first to report on patients with IBD and their perception of the COVID-19 pandemic. Second, our participants came from two large Italian cohorts of patients with IBD who lived in similar epidemic settings. Additionally, more than half

of the eligible patients responded to our survey. Thus, we achieved a larger study population than that of a recent study [34]. Lastly, we conducted this study as soon as the lockdown occurred, and at the time, it was uncertain whether COVID-19 restrictions would be lifted.

Our study also had some limitations. First, half of the patients who received the survey did not answer it, and we cannot rule out the possibility that those who responded were those most worried about the COVID-19 pandemic. Typically, optional online surveys are more likely to be answered by educated and healthy subjects [35]. Therefore, selection bias may be present in our study. Second, we did not have clinical data on IBD activity at the time of the study. As such, we could not assess the relationship between IBD activity and COVID-19 perception. Lastly, we used a standardized ad hoc questionnaire because there was no other available questionnaire in the literature. However, we formulated open questions and multiple-choice answers to avoid influencing the responses of our patients.

Our findings may be useful for future epidemics and may help to establish new ways of managing patients who live outside of hospital areas or are unable to access tertiary outpatient clinics for routine care due to their clinical conditions.

In conclusion, the results of our survey demonstrate that the lockdown had a significant impact on the psychological aspects of our patients and suggest the need for increasing communication with patients with IBD to ensure they receive adequate health care, correct information, and proper psychological support. Our results also indicate that remote medical visits may be well received by young patients with IBD and that IBD centers should implement a remote health care approach. Finally, our survey results highlight the need for encouraging flu vaccinations, as well as future vaccinations for preventing SARS-CoV-2 infection, to the public.

Authors' Contributions

FZ and CC contributed to the study design. FZ conducted all data analyses. MS, EVS, BB, LC, RD, FRD, and SC performed patient recruitment and data collection. FZ and CC wrote the first draft of the paper. FZ, CC, MS, EVS, BB, LC, RD, FRD, and SC wrote, reviewed, and edited the paper.

Conflicts of Interest

None declared.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]

2. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3]

3. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]

4. Geldsetzer P. Knowledge and Perceptions of COVID-19 Among the General Public in the United States and the United Kingdom: A Cross-sectional Online Survey. Ann Intern Med 2020 Jul 21;173(2):157-160. [doi: 10.7326/M20-0912] [Medline: 32196071]

5. Tinsley A, Navabi S, Williams E, Liu G, Kong L, Coates MD, et al. Increased Risk of Influenza and Influenza-Related Complications Among 140,480 Patients With Inflammatory Bowel Disease. Inflamm Bowel Dis 2019 Jan 10;25(2):369-376. [doi: 10.1093/ibd/izy243] [Medline: 30020478]

6. Ford AC, Peyrin-Biroulet L. Opportunistic infections with anti-tumor necrosis factor-α therapy in inflammatory bowel disease: meta-analysis of randomized controlled trials. Am J Gastroenterol 2013 Aug;108(8):1268-1276. [doi: 10.1038/ajg.2013.138] [Medline: 23649185]

7. Osterman MT, Sandborn WJ, Colombel J, Peyrin-Biroulet L, Robinson AM, Zhou Q, et al. Crohn's Disease Activity and Concomitant Immunosuppressants Affect the Risk of Serious and Opportunistic Infections in Patients Treated With Adalimumab. Am J Gastroenterol 2016 Dec;111(12):1806-1815. [doi: 10.1038/ajg.2016.433] [Medline: 27670599]

8. Singh S, Murad MH, Fumery M, Dulai PS, Sandborn WJ. First- and Second-Line Pharmacotherapies for Patients With Moderate to Severely Active Ulcerative Colitis: An Updated Network Meta-Analysis. Clin Gastroenterol Hepatol 2020 Sep;18(10):2179-2191.e6. [doi: 10.1016/j.cgh.2020.01.008] [Medline: 31945470]

9. Wheat CL, Ko CW, Clark-Snustad K, Grembowski D, Thornton TA, Devine B. Inflammatory Bowel Disease (IBD) pharmacotherapy and the risk of serious infection: a systematic review and network meta-analysis. BMC Gastroenterol 2017 Apr 14;17(1):52 [FREE Full text] [doi: 10.1186/s12876-017-0602-0] [Medline: 28407755]

10. Langhorst J, Hofstetter A, Wolfe F, Häuser W. Short-term stress, but not mucosal healing nor depression was predictive for the risk of relapse in patients with ulcerative colitis: a prospective 12-month follow-up study. Inflamm Bowel Dis 2013 Oct;19(11):2380-2386. [doi: 10.1097/MIB.0b013e3182a192ba] [Medline: 23942564]

11. Mardini HE, Kip KE, Wilson JW. Crohn's disease: a two-year prospective study of the association between psychological distress and disease activity. Dig Dis Sci 2004 Mar;49(3):492-497. [doi: 10.1023/b:ddas.0000020509.23162.cc] [Medline: 15139504]

12. D'Amico F, Peyrin-Biroulet L, Danese S. Inflammatory Bowel Diseases and COVID-19: The Invisible Enemy. Gastroenterology 2020 Jun;158(8):2302-2304 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.032] [Medline: 32305331]

13. Savarino EV, Iovino P, Santonicola A, Ghisa M, Laserra G, Barberio B, et al. Clinical and Psychological Impact of COVID-19 Infection in Adult Patients with Eosinophilic Gastrointestinal Disorders during the SARS-CoV-2 Outbreak. J Clin Med 2020 Jun 26;9(6):2011 [FREE Full text] [doi: 10.3390/jcm9062011] [Medline: 32604895]

14. Siniscalchi M, Zingone F, Savarino EV, D'Odorico A, Ciacci C. COVID-19 pandemic perception in adults with celiac disease: an impulse to implement the use of telemedicine. Dig Liver Dis 2020 Oct;52(10):1071-1075 [FREE Full text] [doi: 10.1016/j.dld.2020.05.014] [Medline: 32425731]

15. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

16. Burisch J, Zammit SC, Ellul P, Turcan S, Duricova D, Bortlik M, Epi-IBD group. Disease course of inflammatory bowel disease unclassified in a European population-based inception cohort: An Epi-IBD study. J Gastroenterol Hepatol 2019 Jun;34(6):996-1003. [doi: 10.1111/jgh.14563] [Medline: 30562421]

17. Marinelli CE, Savarino E, Inferrera M, Lorenzon G, Rigo A, Ghisa M, et al. Factors Influencing Disability and Quality of Life during Treatment: A Cross-Sectional Study on IBD Patients. Gastroenterol Res Pract 2019;2019:5354320 [FREE Full text] [doi: 10.1155/2019/5354320] [Medline: 31531015]

18. Marinelli C, Savarino EV, Marsilio I, Lorenzon G, Gavaruzzi T, D'Incà R, et al. Sleep disturbance in Inflammatory Bowel Disease: prevalence and risk factors - A cross-sectional study. Sci Rep 2020 Jan 16;10(1):507 [FREE Full text] [doi: 10.1038/s41598-020-57460-6] [Medline: 31949257]

19. Carpinelli L, Bucci C, Santonicola A, Zingone F, Ciacci C, Iovino P. Anhedonia in irritable bowel syndrome and in inflammatory bowel diseases and its relationship with abdominal pain. Neurogastroenterol Motil 2019 Mar 10;31(3):e13531. [doi: 10.1111/nmo.13531] [Medline: 30628137]

20. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]

21. Rosini U. PCM-DCP dati forniti dal Ministero della Salute. GitHub. URL: https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20200324.pdf [accessed 2020-10-26]

22. Weber J, Javelle F, Klein T, Foitschik T, Crucian B, Schneider S, et al. Neurophysiological, neuropsychological, and cognitive effects of 30 days of isolation. Exp Brain Res 2019 Jun 29;237(6):1563-1573. [doi: 10.1007/s00221-019-05531-0] [Medline: 30927043]

23. Haider II, Tiwana F, Tahir SM. Impact of the COVID-19 Pandemic on Adult Mental Health. Pak J Med Sci 2020 May;36(COVID19-S4):S90-S94 [FREE Full text] [doi: 10.12669/pjms.36.COVID19-S4.2756] [Medline: 32582321]

24. Ciacci C, Siniscalchi M. Tips from the battlefront: Psychological support of patients with a chronic illness during the COVID-19 lockdown in four steps. United European Gastroenterol J 2020 Jul;8(6):741-742 [FREE Full text] [doi: 10.1177/2050640620927564] [Medline: 32423309]

25. Raony Í, de Figueiredo CS, Pandolfo P, Giestal-de-Araujo E, Oliveira-Silva Bomfim P, Savino W. Psycho-Neuroendocrine-Immune Interactions in COVID-19: Potential Impacts on Mental Health. Front Immunol 2020;11:1170 [FREE Full text] [doi: 10.3389/fimmu.2020.01170] [Medline: 32574266]

26. 1st Interview COVID-19ECCO Taskforce. ECCO. 2020 Mar 13. URL: https://www.ecco-ibd.eu/images/6_Publication/6_8_Surveys/1st_interview_COVID-19%20ECCOTaskforce_published.pdf [accessed 2020-10-15]

27. Ungaro RC, Sullivan T, Colombel J, Patel G. What Should Gastroenterologists and Patients Know About COVID-19? Clin Gastroenterol Hepatol 2020 Jun;18(7):1409-1411 [FREE Full text] [doi: 10.1016/j.cgh.2020.03.020] [Medline: 32197957]

28. Zingone F, Savarino EV. Viral screening before initiation of biologics in patients with inflammatory bowel disease during the COVID-19 outbreak. Lancet Gastroenterol Hepatol 2020 Jun;5(6):525 [FREE Full text] [doi: 10.1016/S2468-1253(20)30085-6] [Medline: 32220656]

29. Zingone F, Buda AE, Savarino EV. Starting a biologic therapy in IBD patients amidst COVID-19: hold, careful monitoring or testing? J Crohns Colitis 2020 May 19:jjaa102 [FREE Full text] [doi: 10.1093/ecco-jcc/jjaa102] [Medline: 32426800]

30. Danese S, Cecconi M, Spinelli A. Management of IBD during the COVID-19 outbreak: resetting clinical priorities. Nat Rev Gastroenterol Hepatol 2020 May;17(5):253-255 [FREE Full text] [doi: 10.1038/s41575-020-0294-8] [Medline: 32214232]

31. Farraye FA, Melmed GY, Lichtenstein GR, Kane SV. ACG Clinical Guideline: Preventive Care in Inflammatory Bowel Disease. Am J Gastroenterol 2017 Feb;112(2):241-258. [doi: 10.1038/ajg.2016.537] [Medline: 28071656]

32. Farshidpour M, Charabaty A, Mattar MC. Improving immunization strategies in patients with inflammatory bowel disease. Ann Gastroenterol 2019;32(3):247-256 [FREE Full text] [doi: 10.20524/aog.2019.0351] [Medline: 31040621]

33. Bezzio C, Saibeni S, Variola A, Allocca M, Massari A, Gerardi V, Italian Group for the Study of Inflammatory Bowel Disease (IG-IBD). Outcomes of COVID-19 in 79 patients with IBD in Italy: an IG-IBD study. Gut 2020 Jul;69(7):1213-1217 [FREE Full text] [doi: 10.1136/gutjnl-2020-321411] [Medline: 32354990]

34. Hardigan PC, Popovici I, Carvajal MJ. Response rate, response time, and economic costs of survey research: A randomized trial of practicing pharmacists. Res Social Adm Pharm 2016;12(1):141-148. [doi: 10.1016/j.sapharm.2015.07.003] [Medline: 26329534]

35. Fitzgerald D, Hockey R, Jones M, Mishra G, Waller M, Dobson A. Use of Online or Paper Surveys by Australian Women: Longitudinal Study of Users, Devices, and Cohort Retention. J Med Internet Res 2019 Mar 14;21(3):e10672 [FREE Full text] [doi: 10.2196/10672] [Medline: 30869647]

Abbreviations

IBD: inflammatory bowel disease

XSL•FO

RenderX

Original Paper

Communicative Blame in Online Communication of the COVID-19 Pandemic: Computational Approach of Stigmatizing Cues and Negative Sentiment Gauged With Automated Analytic Techniques

Angela Chang[1,2], PhD; Peter Johannes Schulz[2], PhD; ShengTsung Tu[3], PhD; Matthew Tingchi Liu[4], PhD

[1]Department of Communication, Faculty of Social Sciences, University of Macau, Macao, China

[2]Institute of Communication and Health, Lugano University, Lugano, Switzerland

[3]Department of Radio and Television, Ming Chuan University, Taipei, Taiwan

[4]Department of Management and Marketing, Faculty of Business Administration, University of Macau, Macao, China

Corresponding Author:
Angela Chang, PhD
Department of Communication
Faculty of Social Sciences
University of Macau
E21-2056, Avenida da Universidade, Taipa
Macao, 100
China
Phone: 853 88228991
Email: wychang@um.edu.mo

Abstract

Background: Information about a new coronavirus emerged in 2019 and rapidly spread around the world, gaining significant public attention and attracting negative bias. The use of stigmatizing language for the purpose of blaming sparked a debate.

Objective: This study aims to identify social stigma and negative sentiment toward the blameworthy agents in social communities.

Methods: We enabled a tailored text-mining platform to identify data in their natural settings by retrieving and filtering online sources, and constructed vocabularies and learning word representations from natural language processing for deductive analysis along with the research theme. The data sources comprised of ten news websites, eleven discussion forums, one social network, and two principal media sharing networks in Taiwan. A synthesis of news and social networking analytics was present from December 30, 2019, to March 31, 2020.

Results: We collated over 1.07 million Chinese texts. Almost two-thirds of the texts on COVID-19 came from news services (n=683,887, 63.68%), followed by Facebook (n=297,823, 27.73%), discussion forums (n=62,119, 5.78%), and Instagram and YouTube (n=30,154, 2.81%). Our data showed that online news served as a hotbed for negativity and for driving emotional social posts. Online information regarding COVID-19 associated it with China—and a specific city within China through references to the "Wuhan pneumonia"—potentially encouraging xenophobia. The adoption of this problematic moniker had a high frequency, despite the World Health Organization guideline to avoid biased perceptions and ethnic discrimination. Social stigma is disclosed through negatively valenced responses, which are associated with the most blamed targets.

Conclusions: Our sample is sufficiently representative of a community because it contains a broad range of mainstream online media. Stigmatizing language linked to the COVID-19 pandemic shows a lack of civic responsibility that encourages bias, hostility, and discrimination. Frequently used stigmatizing terms were deemed offensive, and they might have contributed to recent backlashes against China by directing blame and encouraging xenophobia. The implications ranging from health risk communication to stigma mitigation and xenophobia concerns amid the COVID-19 outbreak are emphasized. Understanding the nomenclature and biased terms employed in relation to the COVID-19 outbreak is paramount. We propose solidarity with communication professionals in combating the COVID-19 outbreak and the infodemic. Finding solutions to curb the spread of virus bias, stigma, and discrimination is imperative.

XSL·FO
RenderX

KEYWORDS

placing blame; culprits; sentiment analysis; infodemic analysis; political grievances; COVID-19; communication; pandemic; social media; negativity; infodemic; infodemiology; infoveillance; blame; stigma

Introduction

Background of COVID-19 and Blaming Devices

Toward the end of 2019, a new coronavirus appeared in the city of Wuhan, Hubei Province, mainland China. On February 11, 2020, the World Health Organization (WHO) officially named the new human infectious disease "COVID-19" [1]. On March 11, 2020, it was designated as a global pandemic, spreading across 185 countries and regions [1]. The ongoing COVID-19 pandemic may have been inevitable due to the virus's fast transmission and highly contagious nature. To date, according to the Johns Hopkins University dashboard as of August 8, 2020, there has been an overall worldwide total of 15,751,658 confirmed cases and 639,207 deaths [2]. No one could have imagined COVID-19's rapid global spread and devastating impact.

Governments have been criticized for failing to take adequate action against COVID-19. For instance, the Chinese government was blamed for not controlling the animal trade, which was alleged to have caused the infection in humans. Early discourse contained several contributions suggesting that COVID-19 could have originated from a laboratory in Wuhan [3]. Information was spread in spite of lacking tenable scientific evidence concerning the virus's pathology. US President Donald Trump and his administration harshly blamed China for its failure to contain COVID-19 and, by calling COVID-19 the *Chinese virus*, potentially incited racism and inadvertently attacked people of Asian descent around the world [4]. The discourse was of such low quality that a group of 27 prominent public health scientists from outside of China dismissed the biased information and pushed for a firm condemnation of misinformation and conspiracy theories about the origin and facts surrounding the virus (eg, [5]). These examples show that not only the virus but also the way it is spoken about can hurt people.

Social stigma in the context of a disease outbreak comes from an impulse to assign blame; hence, abundant research has acknowledged the social stigma and the subsequent blame and discrimination attached to COVID-19 (eg, [6-8]). The lack of a clear understanding about social stigma regarding the COVID-19 pandemic may lead to the circulation of false blame and negative bias, which jeopardizes the public's psychosocial development and well-being. As such, it is necessary to address disease-related stigma during infectious disease outbreaks by examining stigmatizing cues and negative sentiments along with blaming information.

As of August 8, 2020, a total of 477 confirmed cases, 7 deaths, and 83,117 tested people were reported by Taiwan's Centers for Disease Control (CDC) [9]. With the outbreak of the deadly COVID-19, Taiwan might have been in for a difficult time because of its close ties with mainland China. However, unlike many neighboring countries and regions, Taiwan has a comparatively low case-fatality rate and has not imposed a strict city lockdown. Instead, people in Taiwan have been urged to reduce their contact with others by maintaining social distancing, washing their hands frequently, and wearing face masks at all times [9]. These requests from government created many social dilemmas and violent altercations, particularly during the COVID-19 crisis. In fact, such information involved social stigma, complaining, and collective blaming often expressed through online communication to form public opinions and, in turn, affect people's cognition. Blame is a vehicle for making meaning, through which the lay public seeks to understand unexpected risky events. Meanwhile, blaming someone is the practice of holding that agent responsible while expressing attitudes of resentment, indignation, or grievance.

Attributing Blame and Stigma

A precondition for blame attribution is the belief that in a just world, where people behave fairly, everyone gets what they deserve [10-12]. As such, three variables including attribution of blame, responsibility, and actor intention were examined to show that the accused agents are in fact the culprits [13,14]. Previous research on blaming attribute of health risks indicates that it leads to a range of damaging social outcomes. For instance, during the COVID-19 outbreak, Chinese people living overseas experienced discrimination, and the majority of Chinese people in China exhibited discriminatory attitudes toward Chinese themselves [15]. When considering various aspects of communication in relation to public health issues, the COVID-19 pandemic tends to provoke xenophobia, discrimination, and biases with stigmatized monikers [16-18]. This shows that blaming discourse and the public's lack of understanding of sensationalized media discourse is intricately connected to the specificities of social conditions. The source of health information for the lay public may deflect and diffuse blame with stigmatizing cues. In particular, ubiquitous online media has created an information overload that makes it difficult to differentiate between true and false information [19-21].

That certain classes of people and agents become targets for receiving blame for negative health outcomes is not a new phenomenon. For instance, placing blame on China is a familiar phenomenon and has contributed to the notion that China is an unsanitary entity. Some reporters held Chinese immigrants in New York's Chinatown responsible for the 2013 severe acute respiratory syndrome (SARS) outbreak, despite public health officials failing to find SARS cases there [22]. The assumed health status of Chinese people led to the stigmatization of that community. Studies have emphasized that public responses and negative attitudes toward certain groups have contributed to the spread of the COVID-19 pandemic [23-25]. The naming of health risks for specific groups can fuel the negative consequences that occur as a result of the fight against infectious diseases. Therefore, researchers have confirmed the widespread collective perceptual bias against the Chinese by using stigmatizing monikers [8].

Blaming can reflect the user's emotional state and efforts at mitigating potential losses. In particular, the use of negative tones in the background of communications can reveal the interlocutors' intent, as they tend to come from existing sentiments of frustration and grievance [26-28]. Thus, assessing the valence of messages from news reports and commentary on COVID-19 as negative or positive can help researchers reach a basic understanding of the context.

When the public is coping with unexpected health-related risk events, stigmatizing monikers can indicate who is giving and receiving blame based on valence [29,30]. Social media has amplified the problem of stigma by spreading inaccurate and harmful information. The harm is not only medical but also includes discrimination against people at the epicenter of an outbreak [16,17,31,32]. A recent study endorsed the problem of stigma by extracting sentiment keywords from Twitter hashtags related to COVID-19 [33]. They found that the keywords "corona" and "Wuhan corona" were associated with emotions of fear and anger, while the least common emotions expressed in tweets were sadness, joy, and disgust. Perceptions of risk can be amplified or attenuated by a variety of emotions including perceived dreadfulness, lack of controllability, and unfamiliarity as projected through all types of media. Social media has proven to be one of the most influential platforms for interacting about controversial topics and making aggressive or contemptuous comments [34-38]. A positive correlation between the number of Weibo posts and the number of reported COVID-19 cases in Wuhan showed that approximately 10 additional cases were reported per 40 Weibo posts [39]. Notably, the effect size was said to be larger in Wuhan than what was observed in other cities in China.

Goal of This Study

To summarize, this study sought to analyze how frequently online media was used to disseminate COVID-19–related information with stigmatizing cues, examine how frequently principal agents in the field are put in proximity to negative sentiment regarding the COVID-19 pandemic, and assess discourse regarding COVID-19 over time by taking into account blaming sentiment. This study draws on existing theories to understand how social stigmas and subsequent blaming present challenges, as nations grapple with restrictions on individuals' movement and move to more normal social interaction.

Research Questions

Four research questions (RQs) were raised to identify the interlocutor's intent and the extent to which online media has attributed blame along with the collective expression of sentiment.

- RQ 1: How much coverage and discussion is devoted to COVID-19 and its related topics in news media and other social media sources?
- RQ 2: What stigmatizing terms are mentioned with sentiment in discussions related to COVID-19?
- RQ 3: Which targets are blamed most for the pandemic in online media?
- RQ 4: What association is there between blaming sentiment and media source in the COVID-19 pandemic?

Methods

Automated Computational Approach

Natural language processing (NLP) in the field of machine learning, which enables a computer to analyze, manipulate, and potentially generate human language, has been widely applied worldwide (eg, [32,35,40]). A machine learning algorithm, DivoMiner in Taiwan, with the ability to automatically identify and classify patterns in large amounts of data sets was employed. This tailored text-mining platform assisted in gaining insights from an unstructured text corpus for key terms, phrases, and sentiment assessment. The collection of various digital communication platforms was converged through automated technology, allowing for real-time aggregation, organization, and analysis of the COVID-19 discourse [17].

The timeline for this study was from December 30, 2019, to March 31, 2020, which ensured the provision of timely information related to COVID-19. It also ensured that the time periods are comparable for operational reasons such as why certain weeks may have higher demand than others or other factors that could influence blaming discourse. The research strategy included enabling DivoMiner to identify data in their natural settings by retrieving and filtering online media sources [41,42] and constructing vocabularies and learning word representations from NLP for analysis along with the research theme (stigmatizing and sentiment language) from online media genres [41-43].

Data Collection

The leading digital news platforms in Taiwan were recruited based on their high use. At the same time, social networks and discussion forums have become increasingly important sources of news. Among all the social networking services, those with the highest penetration were YouTube and Instagram, as they reach approximately 23 million (89%) users, followed by Facebook, which has over 21 million (82%) active users, and open discussion forums, which reach over 3.3 million (95.5%) users aged between 12-38 years [41]. The aforementioned media platforms reach broad segments of the population with a daily flow of health news and discourse. In sum, the media source data recruited in this study comprised of ten mainstream news services, eleven discussion forums, one social networking service (Facebook), and two principal media sharing networks (Instagram and YouTube) for their population data set.

All publicly accessible online communications containing the target keywords posted within the 3-month timeline were automatically collected via the DivoMiner application. For example, a mainstream online newspaper, Apple Daily News, and their Facebook fan page were recruited, and the largest terminal-based bulletin board system, PTT, was observed [41,44]. However, it is worth noting that some popular social media platforms such as Twitter can serve as an accurate mirror of the population in the English-speaking world but not in the Chinese-speaking world. Hence, limiting text data to Chinese-speaking regions serves as a sufficiently effective means of gaining insight. Additionally, issues such as ethical consideration and the legality of subsequent privacy violations were less of a concern.

To ensure the efficiency of capturing opinions related to COVID-19 from unstructured text in syntactically explicit language, data were trained to include some knowledge of semantic meaning in our model [42,43,45]. To verify the feasibility and reliability of the word embedding, three coders manually labeled and checked 1500 random postings. After testing and training, the classification accuracy level reached 75% (1125/1500), which was deemed to be acceptable [41,42]. After irrelevant opinionated data were excluded (eg, shopping, nonnews), the DivoMiner classifier implemented the filtering process of the recorded data in digital form.

Codebook Development

To parse meaning from online texts, all terms and keywords related to COVID-19 were initially collected from the official document issued by the China International Publishing Group in February 2020 [46]. To meet the requirements of computational analysis, the Word2Vec technique was employed to find continuous embedding of words. Word2Vec learns from reading 356,901 articles from the Chinese Wikipedia corpus and memorizing which words tend to appear in similar contexts [43]. After pretraining on a large corpus, it generates a multidimensional vector for each word in a vocabulary, with words of similar meaning being closer to each other. A total of 106 confirmed keywords resulted from several pilot tests creating logical terms and phrases that DivoMiner could assist in analyzing.

An example of searching a taxonomy in the tailored platform uMiner was "肺炎" (pneumonia) and the four alternative terms were "病毒性肺炎" (virus pneumonia), "冠狀病毒" (corona virus novel coronavirus), "新型冠狀肺炎" (novel coronavirus), and "2019新型冠狀病毒" (2019 novel coronavirus). The variables measuring stigmatizing monikers recruited were based on NLP for collective behavioral propensities against China or Chinese people [30,32,45]; five widely used and problematic terms with biases were "武漢病毒" (Wuhan virus), "武漢肺炎" (Wuhan pneumonia), "中國肺炎" (China pneumonia), "中國人肺炎" (Chinese pneumonia), and "中國病毒" (China virus).

DivoMiner contains a module for determining the word sentiment of each opinion. The variable measuring three sentiment tones (positive, neutral, and negative) was based on using the terms "joy," "happy," and "like" as points of reference for a positive tone, while "anger," "fear," and "sadness" were used as points of reference for a negative tone [22,25,26]. The synonyms of emotion words and adjectives not associated with the aforementioned emotions were labelled as having a neutral tone [25-27,30,43]. Negative sentiments are ideal indicators of collective perceptual bias in measuring the blaming of accused culprits [14,19-23]. A randomly assigned group of the 1200 sample posts were cross-checked by the first author and two trained research assistants. Disagreements were resolved by discussion for reaching consensus. An acceptable level of agreement of 77% (924/1200) was reached at the end. The interrater reliability employed the Cohen kappa coefficient by computing sentiment and targeted figures, groups, and organizations for the online postings. The interrater reliability was 0.5901 (95% CI 0.49-0.71; P=.06) and rated as moderate [42,43].

Results

Frequency and Trend of COVID-19 Mentions

We identified a total of 1,073,983 texts about COVID-19 from 24 online sources in Taiwan over 3 months. The first text on COVID-19 was a news article published on HiNet on December 31, 2019. It reported that Taiwan's CDC had started in-flight disinfection and quarantine measures in response to the pneumonia epidemic in Wuhan. Notably, COVID-19 was confirmed to have spread to Taiwan 3 weeks later.

Almost two-thirds of the 1,073,983 texts on COVID-19 came from news services (n=683,887, 63.68%), followed by Facebook (n=297,823, 27.73%), discussion forums (n=62,119, 5.78%), and Instagram and YouTube (n=30,154, 2.81%). The average daily volume of COVID-19 texts was 7354 (range 0-14,561) from news services, 3202 (range 0-5937) from Facebook, 324 (0-678) posts on Instagram and YouTube, and 668 (range 0-1264) on discussion forums.

The amount of COVID-19 coverage and discourse shows several noteworthy patterns. First, until January 19, 2020, the daily number of new stories was always considerably less than 500 but that number rose to more than 1000 on January 20. The news stories exploded on January 21, 2020, with approximately 3500 stories and slightly less than 5500 the day after. From then on, the number of news stories sourced by news services rose to about 12,000 on weekdays and between 6000 and 8000 on weekends. Second, the all-time high was reached in the week beginning on March 16. Four of the five weekdays in that week were days when the intensity of coverage was the greatest over the 3-month time frame. Third, there was a striking weekly rhythm in the time frame, with the number of stories falling from approximately 12,000 stories per day on weekdays to between 6000 and 8000 on weekends. Fourth, the pattern for the other sources by and large mirrors the development of the news services output. Figure 1 provides an overview of the developments.

Figure 1. The intensity of communication about COVID-19 in social media and news services.
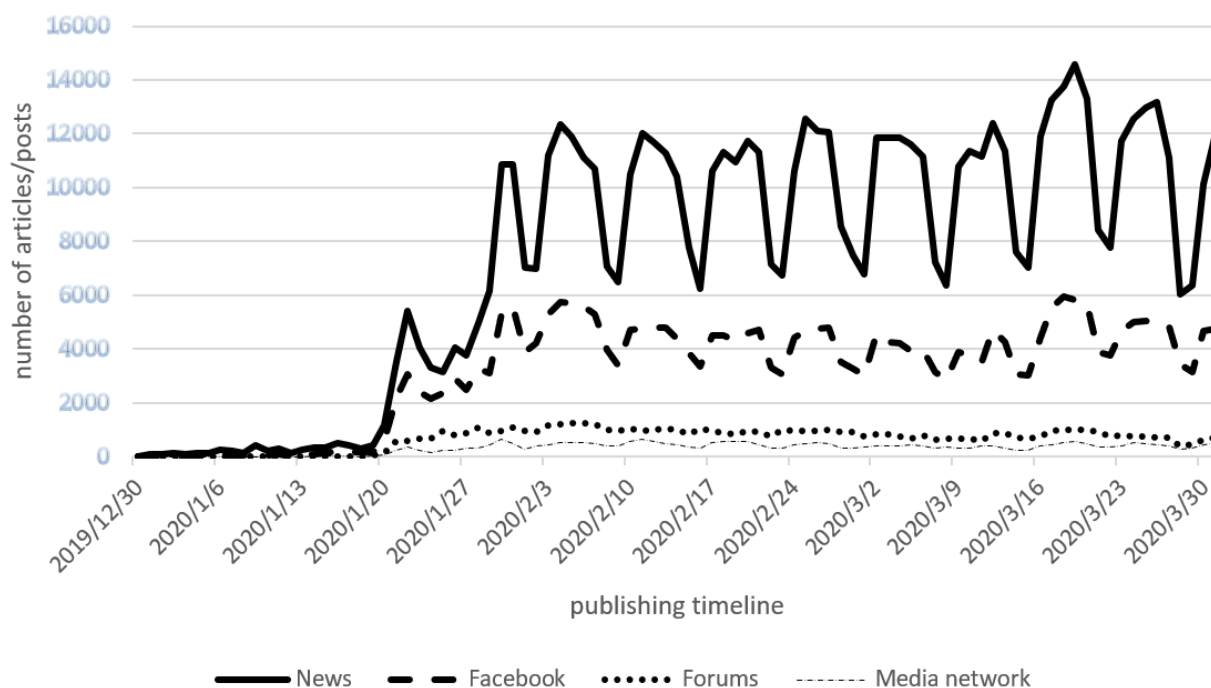
During infectious disease outbreaks, the underlying mechanism of social media posts connecting users' risk perceptions was observed to be frequently high, with resulting cathartic effects. The keyword analysis identified two key phrases, "Tsai Ing-wen" (n=1,082,632; Taiwan's president) and "pneumonia" (n=1,008,486), which received the highest total frequency, followed by the geographical nomenclature "Wuhan" (n=762,004) and "China" (n=343,489). Masks and mask-rationing plans were the fifth most frequently used phrase (n=305,769). Two targets linked to COVID-19 followed suit: ethnic groups of new immigrants (eg, mainland Chinese, Vietnamese, Filipinos, and Indonesians; n=7299) and Dr Li Wenliang, an ophthalmologist who attempted to alert government and the public to the imminent danger in the early phases of the pandemic (n=6004). The results showing a total

of 51 frequent Chinese keywords with English translations is attached in Multimedia Appendix 1. A descriptive analysis of the top 20 high-frequency words in descending order is presented in Table 1.

To investigate the reactions pertaining to each theme, six themes including disease; infection prevention; geographical naming; organizations, institutes, and events; policy; and political figures were frequently covered. In comparison, the two themes that were least covered across various sources over time were those concerning nonpolitical figures and groups and occupations. The mean score gives the measurement of the central tendency for the analysis of thematic data. The descriptive analysis of the attributes of COVID-19 messages are presented in Multimedia Appendix 2 as the total amount of discussion, means, SD, and ranges of key terms assessed.

Table 1. High-frequency words related to COVID-19 in Taiwan's online communication.

COVID-19–related words	Frequency, n
Tsai Ing-Wen (or Taiwan political figure or Taiwan President)	1,082,632
Pneumonia	1,008,486
Wuhan	762,004
Wuhan pneumonia	715,719
China	343,489
Mask or mask-rationing plan	305,769
Coronavirus	261,643
Confirmed case	185,818
Disinfection	99,817
Wash hands often/carefully	96,593
Mandatory quarantine, self-monitored quarantine	91,097
WHO[a]	81,134
Taiwan comrade (or cross-strait charter for Taiwan businessman in China)	69,310
Chen Shih-Chung (or Minister of Health and Welfare)	62,870
City lockdown	51,261
Soo Tsing Tshiong, Prime Minister of Taiwan, Executive Yuan	50,863
Cruise, Westerdam, Aquarius, World Dream, or Diamond Princess	46,906
Xi Jingping or leader of communist party, Chair Xi, or Chair	40,664
Vaccine	29,559
Suspected case	23,926

[a]WHO: World Health Organization.

COVID-19 and Stigmatizing Cues

The usage of "pneumonia" and "virus" can be traced to the day of December 30, 2019. Thus, the stigmatizing term with negative sentiment in discussions related to COVID-19 appeared on the next day. With increasing interest in misusing the term "Wuhan pneumonia" (n=639,456), it comprised approximately one-third of overall media sources. A chi-square test of independence showed that there was a significant association between keyword usage and media platforms (χ^2_9=2,311,455, P<.001). The stigmatizing terms were presented most frequently in the news (n=519,261, 37.0%), followed by Facebook (n=193,249, 34.4%), forums (n=38,650, 36.2%) and other social media networks (Instagram and YouTube; n=19,325, 32.1%). Table 2 presents a comparison of usage between stigmatized and nonstigmatized terms adopted in COVID-19 discussions on different media platforms.

Table 2. Comparison of nonstigma (recommended) keywords and stigmatizing terms by media type.

Variables	Nonstigmatized, n (%)	Stigmatized, n (%)
News (n=1,405,306)	886,045 (63.0)	519,261 (37.0)
Facebook (n=561,298)	368,049 (65.6)	193,249 (34.4)
Forums (n=106,807)	68,157 (63.8)	38,650 (36.2)
Instagram and YouTube (n=60,219)	40,894 (67.9)	19,325 (32.1)
Total (n=2,133,630)	1,363,145 (63.9)	770,485 (36.1)

For the stigmatizing terms with a sentiment assessment, the terms "Wuhan pneumonia" and "China virus," as potentially offensive terms, accumulated a total of 631,192 posts with associated sentiments. A negative tone of more than 50% was associated with blame (n=331,550, 52.3%), compared to a positive tone of 26.86% (n=169,541) and a neutral tone of 20.61% (n=130,101). There was a significant relationship between sentiment and stigmatizing terms (χ^2_9=994,650, P<.001). Table 3 presents a comparison of sentiments associated with stigmatizing terms on different media platforms.

Table 3. Sentiment assessment of stigmatizing terms by media type.

Variables	Positive, n (%)	Neutral, n (%)	Negative, n (%)
News (n=359,574)	106,091 (29.5)	85,681 (23.8)	167,802 (46.7)
Facebook (n=207,768)	49,150 (23.7)	29,084 (14.0)	129,534 (62.3)
Forum (n=43,464)	6866 (15.8)	11,904 (27.4)	24,694 (56.8)
Instagram and YouTube (n=20,386)	7434 (36.5)	3432 (16.8)	9520 (46.7)
Total (n=631,192)	169,541 (26.9)	130,101 (20.6)	331,550 (52.5)

Most Blamed Targets

The targets blamed most often were four leading political figures (Soo Tsing Tshiong, Chen Shih-Chung, Tsai Ing-wen, and Xi Jinping), a group of immigrants to Taiwan, and Dr Li Wenliang in China. The number of sentiment assessments ranged from a minimum of 6004 to a maximum of 935,691. In comparison, the least posts and coverage concerning sentiment were on policy, treatment, and welfare organizations (ranging from 0 to 1328). Quantifying and understanding the development of comments with sentiment in news services and social media activity revealed that a politician, Soo Tsing Tshiong (the Taiwanese Prime Minister), drew the most attention with polarized tones, both negative (n=456,187, 48.8%) and positive (n=280,814, 30.0%). Chen Shih-Chung, the Taiwan Minister of Health and Welfare, followed with more negative (n=22,033, 42.0%) than positive tones (n=15,432, 29.4%), as he oversaw resources across ministries and private stakeholders to fight against COVID-19. Additionally, Xi Jinping, the President of the People's Republic of China (PRC), and issues related to him were associated with more negative (n=8148, 54.6%) than positive tones (n=2859, 19.1%).

Sentiment analysis alongside nonpolitical targets such as new immigrants, foreign labor, and foreign spouses connected more negative (n=3153, 43.2%) than positive emotions (n=2105, 28.8%). Unexpectedly, Li Wenliang, the whistleblower, and issues related to him were associated overwhelmingly with negative (n=4358, 72.6%) than positive discussions (n=390, 6.5%). Dr Li who brought the problem about the impending virus to others' awareness and issues related to him were not praised in the early months of the virus outbreak. Instead, Dr Li and issues related to him became a target of critics. The only exception was Tsai Ing-wen, President of Taiwan, with more

positive comments (n=7786, 45.7%) than negative comments (n=3696, 21.7%) across media. Specifically, the positive comments related with her were mainly in news services (n=5318, 58.0%), compared with social media: Facebook (n=2714, 51.6%), forums (n=704, 39.1%), and YouTube and Instagram (n=252, 34.6%). The sentiment used to discuss this political leader differed considerably across media types (χ^2_9=11,088 P<.001).

Targeting the top six figures and groups, a 3 x 4 analysis of variance, with three sentiment tones (positive, neutral, and negative) and four media platforms (news, Facebook, discussion forums, and Instagram and YouTube) as between-subjects factors, revealed the main effects of tone ($F_{2,60}$=1.13, P=.33) and media platform ($F_{1,60}$=11.90, P=<.001). Hence, post hoc comparisons using the Tukey honestly significant difference test showed a statistically significant difference between the three different media platforms (P=.003): (1) news and Facebook, (2) news and forums, and (3) news and social networks (Instagram and YouTube). The coverage of online news with sentiment tone showed a significantly higher average and SD (mean 4117, SD 3799) than the Facebook posts (mean 1778, SD 1816). The effect sizes for these two significant effects (news and Facebook, and news and forums) were 0.79 and 1.33, respectively. Additionally, sentiment expressed on online forum posts had a significantly higher average score of 530 (SD 346), compared to Instagram and YouTube (average score 225, SD 217), with an effect size of 1.45. Taken together, these results suggest that news coverage with a sentiment tone had an effect on Facebook, forum posts, and Instagram and YouTube. Specifically, our results suggest that when news articles involved emotions, the sentiment carries over to social media. The means and SD for the factorial design are presented in Table 4.

Table 4. Differences between media venue with sentiment on targets according to post hoc tests.

Media[a]	Mean (SD)	Mean differences			
		News	Facebook	Forums	Instagram and YouTube
News	4177 (3799)	__b	—	—	—
Facebook	1778 (1816)	2339 (0.79)[c,d]	—	—	—
Forums	530 (355)	3587 (1.33)[c,e]	12,482	—	—
Instagram and YouTube	225 (217)	3892 (1.45)[c,e]	1553	305	—

[a]Cell size n=18.

[b]Not applicable.

[c]Effect sizes are indicated in parentheses.

[d]$P=.002$.

[e]$P<.001$.

Discussion

Principal Results

This study uncovered a pattern of how the online blame for the COVID-19 pandemic was directed at groups and figures whose influential users had pre-existing grievances with or frustrations about. People's practices of ascribing blame remain fallible, and it seems natural to assume that their perceptual biases are also features of the object of such descriptions and not only features of their very practice. Using social media to follow news about COVID-19 and related topics compensated for traditional news in terms of gathering a diverse and broad variety of general health news. The public develops interpretations of COVID-19 through a variety of resources, most notably representations presented by mainstream online news. Overall, there was a strong and positive correlation between the negative and positive tones of media sources. Increases in discussion frequency were correlated with increases in sentiment tones on online media.

Stigma is disclosed through negatively valenced responses rather than positive ones associated with figures related to COVID-19. For instance, Soo Tsing Tshiong, Taiwan's Prime Minister, was associated with the most blame, despite that Soo is considered to be one of the local figures responsible for prevention and policy implementation. The blame connecting with him appeared mostly in the news (n=519,261, 67.4%), followed by Facebook (n=193,249, 25.1%) and discussion forums (n=38,650, 5.0%). Notably, Xi Jinping, the PRC President, and issues related to him were associated with more negative tone comments from the news (n=4223, 51.8%), followed by Facebook (n=2672, 32.8%) and discussion forums (n=893, 11.0%). The accused agents are in fact the culprits who are associated frequently with negative sentiment.

Taiwan's news media has been described as professional and independent [42]; however, online news services have been facilitating a negative tone. Online news has persistently used stigmatizing terms to initiate the adoption of stigma-related emotion words. In this study, they appeared first in the news and showed a resurgence of use in the week of January 20, 2020; afterwards, they appeared on Facebook and online forums. Online news is a hotbed of negativity and drives negative sentiment and blame in other media. The stigmatizing terms were clearly deemed offensive, and they might have contributed to recent backlashes against China and Chinese people by encouraging and directing blame [6-8]. Understanding the nomenclature and biased terms employed in relation to the COVID-19 outbreak is paramount while considering the online public's responses and feelings around making biased judgments. Stigmatizing language linked to the pandemic used by online media influencers shows a lack of civic responsibility, encouraging bias and hostility.

Limitations

Despite this study's attempts to establish the accuracy of the inferred meaning from all media texts, this study has several limitations related to research design and analytical workflow. First, the online media data set was characterized by a diversity in genres, which did not use fine-grained information. In particular, the data derived from individual social media accounts came in the form of sparse and short texts, which were less likely to lead to insights into the identification of ambiguous information. Working with data from social media remains subjective, and it is challenging to quantify synthesized data that do not necessarily have a closer common claim on objective truth. Second, our automated methods inevitably fell short in reducing a text to a model that encapsulated all important Chinese sentiment lexicons by training sentiment analysis models [43]. In this case, the neutral tag is one of the most important parts of the research problem. Despite preprocessing and postprocessing data being applied to capture the bits of context, tagging criteria should be more consistent. In addition, it should be noted that more efficient blend words or tags with shorter orthographic and phonetic length were often used in social media, compared to the accompanying key terms with formal spelling in traditional news. This difference indicated that data mining based on keyword matching could underestimate actual volume of usage. Last, the current assessment of sentiment tones requires a deeper understanding of frequent user's affective expressions from online media [26]. The classifiers of sentiment assessment might be more precisely generated for measurement by considering the multifaceted nature of Chinese keywords in various media.

XSL•FO

RenderX

Comparison With Prior Work

The WHO and global medical authorities have agreed to veer away from naming illnesses after places or groups of people because using such names could lead to collective perceptual bias, stigma, and inaccurate assumptions [5]. However, consistent with previous studies (eg, [7,8]), the adoption of the problematic moniker "Wuhan pneumonia" had a high frequency of collective production and consumption. Our data also showed the high association of COVID-19 with China and a specific city within China through references like "Wuhan pneumonia," potentially encouraging xenophobia. Comparing the results with earlier findings (eg, [18,38]), news and social networks were observed to be rough proximations of beliefs providing georeferenced sentiment connecting the virtual and material spaces of a health crisis. Additionally, Tsai Ing-wen has been associated with more favorable affective responses than others, while Chen Shih-Chung was later observed to connect to nonblameworthy statements amid the COVID-19 flare-up at the end of March 2020. The xenophobia that spread during the COVID-19 outbreak showed that individuals can enhance the media's exposure and credibility, and consequently, shape the public's views and appraisals. The positive mood toward the two Taiwanese politicians offers an opportunity to reflect on the lessons learned in this pandemic's framing of online heroization dynamics [23].

Conclusions

This study explores the mechanisms of how blame was associated with various targets in online communications during the COVID-19 outbreak in a widely used language other than English. Given the impact of the online discourse about COVID-19 to date, it is crucial to reduce stigma amid the pandemic. This timely report can be used to inform policies and to stimulate research related to how societies deal with pandemics with stigma mitigation. Particularly, sentiment analysis has great potential in tracing sources for predicting the spread of infectious diseases with emotions [26,29,38,45]. Online data sources such as mobile phones can help researchers discover new pathogens at the community level and can be used to leverage big data and intelligent analytics for public health [30].

This study investigating the contentious and distorted nature of online media dynamics concludes that the collective behavior of perceptual bias against COVID-19 existed in daily communications among Taiwanese users. At the local scale, social media users broadly occupy the same geographical turf, which is why it is considered appropriate to explore the pandemic as a reference for future study. Media users have fueled the unprecedented dissemination of stigmatizing terms with negative tones to direct hostility and blame. Because of this, harmful language can have higher stakes, and the risk of offline harm can become exacerbated. Thus, the awareness of blaming devices is promoted through empowering individuals, health communication researchers, health care professionals, and policy makers to take responsibility for their actions. We propose solidarity with communication professionals in combating the COVID-19 outbreak and for finding solutions to curb the spread of virus bias, stigma, and discrimination.

Authors' Contributions

AC and PJS designed the main concepts of this work. AC and ST performed the data collection. ST and MTL were involved with the data interpretation. AC and PJS conducted the data analysis and wrote and edited this paper. AC, PJS, and MTL reviewed and promoted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Categorized keywords in Chinese and translation in English.
[DOCX File , 16 KB - jmir_v22i11e21504_app1.docx]

Multimedia Appendix 2
Topic related to COVID-19 online communication in Taiwan.
[DOCX File , 13 KB - jmir_v22i11e21504_app2.docx]

References

1. Di Gennaro F, Pizzol D, Marotta C, Antunes M, Racalbuto V, Veronese N, et al. Coronavirus diseases (COVID-19) current status and future perspectives: a narrative review. Int J Environ Res Public Health 2020 Apr 14;17(8):2690 [FREE Full text] [doi: 10.3390/ijerph17082690] [Medline: 32295188]

XSL•FO
RenderX

2. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-08-08]

3. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280. [doi: 10.1111/tmi.13383] [Medline: 32052514]

4. Ang YY. When COVID-19 meets centralized, personalized power. Nat Hum Behav 2020 May;4(5):445-447. [doi: 10.1038/s41562-020-0872-3] [Medline: 32273583]

5. Calisher C, Carroll D, Colwell R, Corley RB, Daszak P, Drosten C, et al. Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19. Lancet 2020 Mar 07;395(10226):e42-e43 [FREE Full text] [doi: 10.1016/S0140-6736(20)30418-9] [Medline: 32087122]

6. Nguyen A, Catalan-Matamoros D. Digital mis/disinformation and public engagment with health and science controversies: fresh perspectives from COVID-19. Media Commun 2020 Jun 25;8(2):323-328 [FREE Full text] [doi: 10.17645/mac.v8i2.3352]

7. Dhanani LY, Franz B. Unexpected public health consequences of the COVID-19 pandemic: a national survey examining anti-Asian attitudes in the USA. Int J Public Health 2020 Jul;65(6):747-754 [FREE Full text] [doi: 10.1007/s00038-020-01440-0] [Medline: 32728852]

8. Jo W, Lee J, Park J, Kim Y. Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis. J Med Internet Res 2020 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]

9. COVID-19 (SARS-CoV-2 infection). Taiwan Centers for Diseases Control. URL: https://www.cdc.gov.tw/En [accessed 2020-07-25]

10. Tognazzini NA. Blameworthiness and the affective account of blame. Philosophia 2013 Mar 24;41(4):1299-1312 [FREE Full text] [doi: 10.1007/s11406-013-9428-3]

11. Alicke M. Blaming badly. J Cogn Cult 2008 Jan 01;8(1-2):179-186 [FREE Full text] [doi: 10.1163/156770908x289279]

12. Furnham A. Belief in a just world: research progress over the past decade. Pers Individ Dif 2003 Apr;34(5):795-817 [FREE Full text] [doi: 10.1016/s0191-8869(02)00072-7]

13. Dawtry R, Callan M, Harvey A, Gheorghiu A. Victims, vignettes, and videos: meta-analytic and experimental evidence that emotional impact enhances the derogation of innocent victims. Pers Soc Psychol Rev 2020 Aug;24(3):233-259 [FREE Full text] [doi: 10.1177/1088868320914208] [Medline: 32321373]

14. Khatri N, Brown GD, Hicks LL. From a blame culture to a just culture in health care. Health Care Manage Rev 2009;34(4):312-322. [doi: 10.1097/HMR.0b013e3181a3b709] [Medline: 19858916]

15. Gao Z, Bischoping K. The emergence of an elder-blaming discourse in twenty-first century China. J Cross Cult Gerontol 2018 Jun;33(2):197-215. [doi: 10.1007/s10823-018-9347-7] [Medline: 29732523]

16. Karalis Noel T. Conflating culture with COVID-19: cenophobic repercussions of a global pandemic. Soc Sci Humanities Open 2020;2(1):100044 [FREE Full text] [doi: 10.1016/j.ssaho.2020.100044]

17. He J, He L, Zhou W, Nie X, He M. Discrimination and social exclusion in the outbreak of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2933 [FREE Full text] [doi: 10.3390/ijerph17082933] [Medline: 32340349]

18. Stephens M. A geospatial infodemic: mapping Twitter conspiracy theories of COVID-19. Dialogues Hum Geogr 2020 Jun 23;10(2):276-281 [FREE Full text] [doi: 10.1177/2043820620935683]

19. Gilmore J, Meeks L, Domke D. Why do (we think) they hate us: anti-Americanism, patriotic messages, and attributions of blame. Int J Commun 2013;7:701-721 [FREE Full text]

20. Holton A, Weberling B, Clarke CE, Smith MJ. The blame frame: media attribution of culpability about the MMR-autism vaccination scare. Health Commun 2012;27(7):690-701. [doi: 10.1080/10410236.2011.633158] [Medline: 22236220]

21. Knobloch-Westerwick S, Taylor LD. The blame game: elements of causal attribution and its impact on siding with agents in the news. Commun Res 2008 Oct 07;35(6):723-744 [FREE Full text] [doi: 10.1177/0093650208324266]

22. Eichelberger L. SARS and New York's Chinatown: the politics of risk and blame during an epidemic of fear. Soc Sci Med 2007 Sep;65(6):1284-1295 [FREE Full text] [doi: 10.1016/j.socscimed.2007.04.022] [Medline: 17544189]

23. Atlani-Duault L, Ward JK, Roy M, Morin C, Wilson A. Tracking online heroisation and blame in epidemics. Lancet Public Health 2020 Mar;5(3):e137-e138 [FREE Full text] [doi: 10.1016/S2468-2667(20)30033-5] [Medline: 32085818]

24. Logie CH, Turan JM. How do we balance tensions between COVID-19 public health responses and stigma mitigation? Learning from HIV research. AIDS Behav 2020 Jul;24(7):2003-2006 [FREE Full text] [doi: 10.1007/s10461-020-02856-8] [Medline: 32266502]

25. Roy M, Moreau N, Rousseau C, Mercier A, Wilson A, Atlani-Duault L. Ebola and localized blame on social media: analysis of Twitter and Facebook conversations during the 2014–2015 Ebola epidemic. Cult Med Psychiatry 2020 Mar;44(1):56-79 [FREE Full text] [doi: 10.1007/s11013-019-09635-8] [Medline: 31214902]

26. Gaspar R, Pedro C, Panagiotopoulos P, Seibt B. Beyond positive or negative: qualitative sentiment analysis of social media reactions to unexpected stressful events. Comput Human Behav 2016 Mar 12;56:179-191 [FREE Full text] [doi: 10.1016/j.chb.2015.11.040]

27. Kumar A, Garg G. Systematic literature review on context-based sentiment analysis in social multimedia. Multimed Tools Appl 2019 Feb 23;79(21-22):15349-15380 [FREE Full text] [doi: 10.1007/s11042-019-7346-5]

28. Anagondahalli D, Turner M. Predicting psychological ripple effects: the role of cultural identity, in-group/out-group identification, and attributions of blame in crisis communication. Risk Anal 2012 Apr;32(4):695-712. [doi: 10.1111/j.1539-6924.2011.01727.x] [Medline: 22050442]

29. Chang PF, Bazarova NN. Managing stigma: disclosure-response communication patterns in pro-anorexic websites. Health Commun 2016;31(2):217-229. [doi: 10.1080/10410236.2014.946218] [Medline: 26266693]

30. Oh S, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: the mediating role of self-relevant emotions and public risk perception. Health Commun 2020 Feb 16:1-10. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]

31. Vaezi A, Javanmard SH. Infodemic and risk communication in the era of CoV-19. Adv Biomed Res 2020;9:10 [FREE Full text] [doi: 10.4103/abr.abr_47_20] [Medline: 32309248]

32. Hua J, Shaw R. Corona virus (Covid-19) "infodemic" and emerging issues through a data lens: the case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]

33. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "infodemic": leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. Open Forum Infect Dis 2020 Jul;7(7):ofaa258 [FREE Full text] [doi: 10.1093/ofid/ofaa258] [Medline: 33117854]

34. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]

35. Chua CEH, Storey VC, Li X, Kaul M. Developing insights from social media using semantic lexical chains to mine short text structures. Decis Support Syst 2019 Dec 26;127:113142 [FREE Full text] [doi: 10.1016/j.dss.2019.113142]

36. Whiting J, Olufuwote RD, Cravens-Pickens JD, Witting AB. Online blaming and intimate partner violence: a content analysis of social media comments. Qual Rep 2019 Jan 13;24(1):78-94 [FREE Full text]

37. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: how health organizations address the health crisis in their social media engagement. Public Relat Rev 2017 Sep 05;43(3):477-486 [FREE Full text] [doi: 10.1016/j.pubrev.2017.04.009]

38. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]

39. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]

40. van der Meer TGLA. Automated content analysis and crisis communication research. Public Relat Rev 2016 Dec 22;42(5):952-961 [FREE Full text] [doi: 10.1016/j.pubrev.2016.09.001]

41. Chang A, Jiao W. Predicting health communication patterns in follower-influencer networks: the case of Taiwan amid COVID-19. Asian J Public Opin Res 2020 Aug 31;8(3):246-264 [FREE Full text] [doi: 10.15206/ajpor.2020.8.3.246]

42. Chang A, Schulz PJ, Cheong AW. Online newspaper framing of non-communicable diseases: comparison of Mainland China, Taiwan, Hong Kong and Macao. Int J Environ Res Public Health 2020 Aug 03;17(15):5593 [FREE Full text] [doi: 10.3390/ijerph17155593] [Medline: 32756457]

43. Muhammad A, Wiratunga N, Lothian R. Contextual sentiment analysis for social media genres. Knowledge-Based Syst 2016 Sep 15;108:92-101 [FREE Full text] [doi: 10.1016/j.knosys.2016.05.032]

44. Statista. Penetration of leading social networks in Taiwan as of 3rd quarter 2019. 2020 Feb. URL: https://www.statista.com/statistics/295611/taiwan-social-network-penetration/ [accessed 2020-10-01]

45. Chen Z, Hu K. Radical enhanced Chinese word embedding. In: Sun M, Liu T, Wang X, Liu Z, Liu Y, editors. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Cham: Springer; Oct 07, 2018:3-11.

46. COVID-19 prevention and control related vocabulary, how to say in English? China Academy of Translation. 2020 Feb 05. URL: http://beta.tac-online.org.cn/index.php?m=content&c=index&a=show&catid=395&id=3611 [accessed 2020-10-01]

Abbreviations

CDC: Centers for Disease Control
NLP: natural language processing
PRC: People's Republic of China
RQ: research question
SARS: severe acute respiratory syndrome
WHO: World Health Organization

XSL•FO

RenderX

Original Paper

Health Perceptions and Misconceptions Regarding COVID-19 in China: Online Survey Study

Jiawei Zhou[1], MPA; Bishwajit Ghose[2], PhD; Ruoxi Wang[1], PhD; Ruijun Wu[3], PhD; Zhifei Li[3], PhD; Rui Huang[2], PhD; Da Feng[2], PhD; Zhanchun Feng[1], PhD; Shangfeng Tang[1], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]School of Pharmacy, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3]China National Center for Biotechnology Development, Beijng, China

Corresponding Author:
Shangfeng Tang, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan
China
Phone: 86 13349895639
Email: sftang2018@hust.edu.cn

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/7/e39493

Abstract

Background: Great efforts have been made to prevent the spread of COVID-19, including national initiatives to promote the change of personal behaviors. The lessons learned from the 2003 SARS outbreak indicate that knowledge and attitudes about infectious diseases are related to panic among the population, which may further complicate efforts to prevent the spread of infectious diseases. Misunderstandings may result in behaviors such as underestimation, panic, and taking ineffective measures to avoid infection; these behaviors are likely to cause the epidemic to spread further.

Objective: The purpose of this study is to assess public health perceptions and misunderstandings about COVID-19 in China, and to propose targeted response measures based on the findings to control the development of the epidemic.

Methods: The study was conducted in April 2020 through an online survey, with participants in 8 provinces in Eastern, Central, and Western China. We designed a questionnaire with a health knowledge section consisting of 5 questions (4 conventional questions and 1 misleading question) on clinical features of and preventive measures against COVID-19. Descriptive statistics, chi-square analysis, binary logistic regression, and Mantel-Haenszel hierarchical analysis were used for statistical analysis.

Results: In total, 4788 participants completed the survey and the mean knowledge score was 4.63 (SD 0.67), gained mainly through experts (76.1%), television (60.0%), newspapers (57.9%), and opinions (46.6%) and videos (42.9%) from social media. Compared to those who obtained information from only 1 or 2 channels, people who obtained information from >3 channels had increased health perception and a better ability to identify misleading information. Suggestions from experts were the most positive information source ($\chi2$=41.61), while information on social media was the most misleading. Those aged >60 years (OR 1.52, 95% CI 1.10-2.11), those with a lower or middle income (OR 1.36, 95% CI 1.00-1.83), those not working and not able to work (OR 1.83, 95% CI 1.04-3.21), those with a household income <100,000 RMB (<US $14,954; OR 1.34, 95% CI 1.08-1.67), and those with >2 suspected symptoms (OR 2.95, 95% CI 1.50-5.80) were more likely to be misled by videos on social media, but the error correction effect of expert advice was limited in these groups.

Conclusions: Multiple information channels can improve public health perception and the identification of misleading information during the COVID-19 pandemic. Videos on social media increased the risk of rumor propagation among vulnerable groups. We suggest the government should strengthen social media regulation and increase experts' influence on the targeted vulnerable populations to reduce the risk of rumors spreading.

XSL•FO
RenderX

Introduction

In December 2019, the novel coronavirus SARS-CoV-2 caused an outbreak of COVID-19 in Wuhan, China [1,2]. COVID-19 is spread by human-to-human transmission via droplets or direct contact [3-5], and the clinical symptoms of COVID-19 include fever, cough, fatigue, and gastrointestinal infection symptoms [6-8]. Major transmission hotspots were brought under control in China, but others subsequently sprouted worldwide. Since late February 2020, the daily number of new cases has been higher in other regions of the world [9]. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic [10], and by May 17, 2020, there were more than 4,500,000 confirmed cases and over 300,000 reported deaths from COVID-19 worldwide [11]. The COVID-19 pandemic has clearly entered a new stage, with rapid spread in countries outside China. All members of society have to understand and practice measures for self-protection and the prevention of transmission [12].

Individual behavior is crucial for controlling the transmission of COVID-19. Although there have been satisfactory results regarding the safety, tolerability, and immunogenicity of a COVID-19 vaccine [13], a vaccine still faces many failure risks, and producing and administering vaccines to millions of people worldwide takes time. In China, the epidemic situation was brought under control by the implementation of multifaceted public health measures including but not limited to intensive intracity and intercity traffic restriction, social distancing measures, home confinement and centralized quarantine, and improvement of medical resources [5,14,15]. In Western democracies, personal action, rather than government action, might be the most important issue. Early self-isolation, seeking medical advice remotely unless symptoms are severe, and social distancing are key [5]. Strong infection control measures are the primary intervention for minimizing the spread of the virus in both health care settings and the community [16,17]. According to the Knowledge-Attitude-Belief (KAP) theory, knowledge is the basis of behavior change, and belief and attitude are the driving forces of behavior change. Therefore, public perception of dealing with highly infectious respiratory diseases plays a vital role in limiting the spread of the infection [18,19].

The lessons learned from the 2003 SARS outbreak suggest that knowledge and attitudes about infectious diseases are associated with the level of panic emotion among the population, which could further complicate efforts to prevent the spread of a disease [20,21]. Behaviors like underestimation, panic, and taking ineffective measures to avoid infection may affect the fight against COVID-19 [22]. Therefore, understanding what the general public knows regarding COVID-19 and which misperceptions they hold about the condition is important for public health authorities in China and other countries, who aim to design effective information campaigns for epidemic prevention and control [23]. For this study, we conducted a rapid online survey in April 2020 to investigate health perceptions and misperceptions regarding COVID-19 among the general population in China and to identify vulnerable populations and channels being used to spread misinformation.

Methods

Study Design

This study used directional stratified convenient sampling to select residents in Eastern, Central, and Western China. To ensure sufficient representativeness, we included not only provinces with severe epidemics, but also provinces with relatively mild epidemics. Therefore, two provinces with severe epidemics and one province with fewer epidemics were selected in the Eastern, Central, and Western regions at the time of sampling. In April 2020, as the overall epidemic situations in the central and eastern regions were relatively severe, we defined severe epidemics as >1000 cases and less severe epidemics as <500 cases. In addition, since the overall epidemic situation in the western region was relatively less serious, provinces with >500 cases were regarded as relatively severe, and provinces with <200 cases were regarded as relatively mild. Hubei, Hunan, and Shanxi provinces were selected in Central China. In Eastern China, Guangdong, Zhejiang, and Fujian provinces were selected. Due to similar local conditions and customs in Sichuan and Chongqing, we only selected one of the two provinces with relatively higher prevalence of COVID-19 in Western China; Sichuan and Gansu provinces were included. Based on cities' populations and economic influence in the provinces, the provincial capital city and another city were purposely selected for each province. In each city, both urban and rural households were selected, for a total of 60 households. All household residents aged ≥10 years were included in this survey. In total, 7118 residents from 1920 households in 8 provinces (16 cities) were included in this investigation. Due to the lower responsiveness of residents in Guangdong (Guangzhou and the Pearl River delta) and Zhejiang provinces (Hangzhou and Jinhua)—only half of the expected households completed the survey—we combined the two provinces for analysis.

According to the COVID-19 clinical and community management guidelines issued by the national health commission of the People's Republic of China [24,25], we designed a questionnaire to determine residents' health perceptions regarding COVID-19.

We designed a self-administered questionnaire containing 168 questions based on literature research and past experience, including questions about demographic information, physical condition, nutrition and prevention behaviors, knowledge of drug preparation and response, and knowledge level of COVID-19, personal health and risk protection, and psychological stress (Cronbach a=.782). There were 5 questions

XSL•FO

RenderX

in the questionnaire designed to examine the participants' understanding of COVID-19:

1. Washing your hands and wearing a mask frequently could help prevent COVID-19.
2. When people with COVID-19 sneeze or cough around you, is it easier to be infected?
3. Eating a lot of garlic could help prevent COVID-19.
4. Improving your immunity could help fight COVID-19.
5. If infected, older adults have the highest risk of mortality.

The above list consists of four measures to prevent COVID-19 (Q1, Q2, Q4, and Q5) and one misleading measure (Q3), all of which were answered on a right/wrong basis. Each correct answer counts for 1 point, while incorrect answers get no points. The total health perception score ranged from 0 to 5; the higher the score, the higher the knowledge level of COVID-19. Based on the above questions, we defined participants who correctly answered 3 or more of the 5 questions as "perception level above average," and calculated the proportion of participants who answered incorrectly on the misleading question (Q3) to get the "misled" rate.

Data Collection

This investigation was carried out from April 4 to 15, 2020. One project manager in each province was recruited to coordinate the provincial survey and organize the investigation training, and 6 local investigators were recruited based on the household income for each city to send an online questionnaire and control the investigation quality. Half of them were from rural areas and the majority were undergraduates. After receiving data collection training online, every investigator was required to directly send the online questionnaire to 20 local families in their social network, including relatives, friends, and classmates. Every eligible participant in the family was invited to fill out the online questionnaire, which they completed within an average of 15 minutes. One secret gift was sent to encourage the participants to complete the questionnaire. Due to the constraints of their own age and educational level, some older adults are unable to participate in online surveys. For this group of people, we recommended that relatives living with them obtain their answers through oral inquiry and fill out the survey according to their choices.

If it was difficult to survey 20 families connected to the investigators, a supplementary survey was conducted by other investigators to complete the household investigation.

During data collection, we used the following quality control measures:

1. We first conducted a preliminary survey to improve our questionnaire, and then grouped the investigators and trained them on how to perform the investigation.
2. Every investigator was independent.
3. Before sending the online questionnaire, the 20 households and eligible family members were asked to generate a unique questionnaire number. Those aged >60 years accounted for >15% of the participants.

4. Every family's questionnaire was sent one by one by the investigators, who delivered a message that "the one who fills the questionnaire carefully, they will receive one secret gift." Trap questions were included to identify those who did not answer questions carefully.
5. The project manager checked the quality of every questionnaire (based on the survey time threshold value of >450 seconds) as well as the consistency of responses to two group questions.

The protocol was reviewed and ethical approval was granted by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (#2020S107). Oral informed consent was obtained from each participant during the online investigation. A total of 6253 residents aged ≥10 years completed the investigation, and 4788 of them were eligible. The response rate was 87.85% (6253/7118), and the valid response rate was 67.13% (4778/7118).

Statistical Analysis

Descriptive statistical methods were used to summarize data on sociodemographic characteristics and responses to questions concerning knowledge, perceptions, and attitudes toward COVID-19. Data were summarized as frequencies (n) and percentages (%) for categorical variables. The differences in health perception among different groups and the guiding effect of different information sources on people's health perceptions were tested using the chi-square test. Binary logistic regression analyses were employed to identify the potential factors and vulnerable populations related to COVID-19 knowledge. The demographic variables and information sources were set as independent variables, and health perception and the "a lot of garlic helps prevent COVID-19" question were respectively set as the outcome variables. To further identify vulnerable groups (ie, those with a high risk of being misled by the main sources of rumor propagation), a Mantel-Haenszel hierarchical analysis of misleading factors was conducted across subpopulations. Unstandardized regression coefficients (β) and odds ratios (ORs) and their 95% CIs were used to quantify the associations between variables and misconceptions regarding COVID-19. Data analyses were performed using Statistical Package for the Social Sciences (SPSS) software (Version 19.0; IBM Corp). $P<.05$ was considered statistically significant.

Results

Sample Characteristics

The average age of the 4788 participants was 40.9 years (SD 18.51 years, range 10-93 years), 53.0% (n=2540) of them were women, and older adults (>60 years) accounted for 17.0% (n=814) of the total sample. Furthermore, 1155 (24.1%) participants were current students and half of them had a bachelor's degree or higher. In addition, 2191 (45.8%) participants were from Central China, 64.0% (n=3065) of them lived in urban areas, and only 8.7% (n=418) of participants lived alone during the COVID-19 epidemic. Detailed demographic characteristics data are shown in Table 1.

Table 1. Demographic characteristics of participants and knowledge level of COVID-19 by demographic variables.

Characteristics	Participants, n (%)	Perception above average level, n (%)	Misled rates, n (%)
Age (years)		*P*<.001	*P*<.001
≤20	599 (12.5)	563 (94.0)	65 (10.9)
21-40	1774 (37.1)	1704 (96.1)	222 (12.5)
41-60	1601 (33.4)	1521 (95.0)	329 (20.5)
>60	814 (17.0)	733 (90.0)	193 (23.7)
Gender		*P*=.06	*P*=.25
Male	2248 (47.0)	2108 (93.8)	365 (16.2)
Female	2540 (53.0)	2413 (95.0)	444 (17.5)
Marital status		*P*<.001	*P*<.001
Unmarried	1725 (36.0)	1654 (95.9)	189 (11.0)
Married or remarried	2851 (59.5)	2682 (94.1)	566 (19.9)
Divorced or widowed, not remarried	212 (4.4)	185 (87.3)	54 (25.5)
Occupation		*P*<.001	*P*<.001
Seeking employment	300 (6.3)	274 (91.3)	62 (20.7)
Not working (not able to work)	273 (5.7)	239 (87.5)	65 (23.8)
Self-employed shop owner or entrepreneur	569 (11.9)	533 (93.7)	118 (20.7)
Staff member in a government or public institution	615 (12.8)	599 (97.4)	110 (19.7)
Farmer, fisherman, or herdsman	321 (6.7)	287 (89.4)	83 (25.9)
Retired	499 (10.4)	469 (94.0)	112 (22.4)
Student	1155 (24.1)	1109 (96.0)	102 (8.8)
Staff member in a big company	276 (5.8)	264 (95.7)	45 (16.3)
Staff member in a small or medium company	426 (8.9)	410 (96.2)	58 (13.6)
Other	354 (7.4)	337 (95.2)	54 (15.3)
Education (years)		*P*<.001	*P*<.001
≤6	698 (14.6)	621 (89.0)	160 (22.9)
7-9	809 (16.9)	758 (93.7)	166 (20.5)
10-12	865 (18.1)	811 (93.8)	178 (20.6)
13-16	2145 (44.8)	2064 (96.2)	286 (13.3)
>16	271 (5.7)	267 (98.5)	19 (7.0)
Areas		*P*=.09	*P*=.88
Eastern China	1317 (27.5)	1259 (95.6)	227 (17.2)
Central China	2191 (45.8)	2058 (93.9)	371 (16.9)
Western China	1280 (26.7)	1204 (94.1)	211 (16.5)
Type of area		*P*<.001	*P*=.93
Urban	3065 (64.0)	2928 (95.5)	519 (16.9)
Rural	1723 (36.0)	1593 (92.5)	290 (16.8)
Household composition		*P*=.005	*P*=.36
Living with others	4370 (91.3)	4139 (94.7)	745 (17.0)
Living alone	418 (8.7)	382 (91.4)	64 (15.3)
Relative self-reported individual income		*P*<.001	*P*=.03
Low (0%-20%)	1505 (31.4)	1393 (92.6)	266 (17.7)
Low and middle (20%-40%)	1141 (23.8)	1069 (93.7)	214 (18.8)

Characteristics	Participants, n (%)	Perception above average level, n (%)	Misled rates, n (%)
Average (40%-60%)	1816 (37.9)	1743 (96.0)	269 (14.8)
Upper middle (60%-80%)	285 (6.0)	277 (97.2)	55 (19.3)
High (80%-100%)	41 (0.9)	39 (95.1)	5 (12.2)
Household income in 2019, RMB (US $)		$P<.001$	$P=.008$
<100,000 (<14,954)	2074 (43.3)	1920 (92.6)	395 (19.0)
100,000-200,000 (14,954-29,909)	1735 (36.2)	1667 (96.1)	270 (15.6)
200,000-300,000 (29,909-44,864)	579 (12.1)	556 (96.0)	89 (15.4)
300,000-400,000 (44,864-59,819)	193 (4.0)	182 (94.3)	23 (11.9)
>400,000 (59,819)	207 (4.3)	196 (94.7)	32 (15.5)

Health Perception and Attitude Toward COVID-19

The average health perception score was 4.63 (SD 0.67). The number of perception scores above the average level significantly differed across age groups, categories of marital status, occupational categories, education levels, residence area type, individual income level, and household income in 2019 ($P<.001$). In contrast, significant differences in "misled" rates were only found for age groups, categories of marital status, occupational categories, and education levels ($P<.001$; Table 1).

The overall accuracy per question is shown in Table 2. This study showed that the sample residents had a high level of knowledge of preventive measures against COVID-19, but the accuracy of responses to "Eating a lot of garlic could help to prevent COVID-19" and "If infected, older adults have the highest risk of mortality" was low in comparison to the other questions, suggesting that further education efforts are needed to improve residents' ability to identify rumors and their understanding of the susceptibility of specific groups.

Table 2. Questionnaire to examine participants' level of knowledge of COVID-19.

Knowledge	Accuracy (%)
1. Washing your hands and wearing a mask frequently could help to prevent COVID-19.	98.3
2. When people with COVID-19 sneeze or cough around you, is it easier to be infected?	97.0
3. Eating a lot of garlic could help to prevent COVID-19.	83.1
4. Improving your immunity could help you fight COVID-19.	95.5
5. If infected, older adults have the highest risk of mortality.	88.8

The vast majority of the participants held an optimistic attitude toward the COVID-19 epidemic. In total, 83.2% of them believed that the Centers for Disease Control could do better in controlling the risk of recurrence, 84.3% thought that hospitals could do better in controlling the risk of recurrence, and only 20.2% thought that COVID-19 outbreaks could happen again (Figure 1), which is consistent with the high level of awareness of COVID-19 prevention measures among the participants. Accurate guidance from the Chinese government during the epidemic period enabled Chinese citizens to have a good understanding of COVID-19 and avoid unnecessary panic and confusion, which also increased people's confidence in the success of the fight against the epidemic [22].

Figure 1. Participants' level of confidence in the success of the fight against the COVID-19 epidemic.
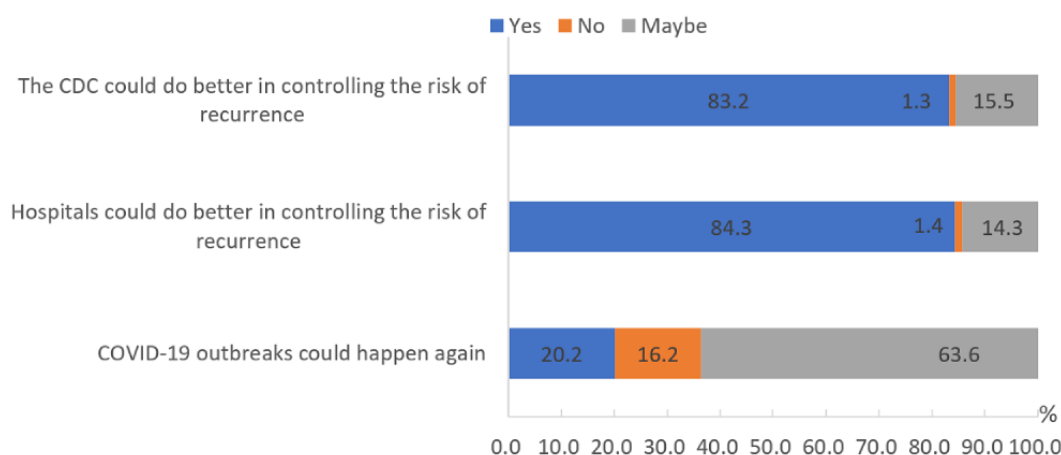
Effect of Various COVID-19 Information Sources

Residents received COVID-19 information through a variety of channels, among which the majority of residents chose mainstream media such as suggestions from experts (76.1%), television (60.0%), and newspapers (57.9%). In addition, it is worth noting that with the advent of the Web 2.0 era and the rise of social media, opinions (46.6%) and videos (42.9%) on social media have also played an important role in this information dissemination. The chi-square test results of answers to the question "Eating a lot of garlic could prevent COVID-19" showed that whether participants obtained information from experts and newspapers is significantly related to whether residents were misled; in particular, expert opinions play a decisive role. During this outbreak, Chinese experts in relevant fields have contributed to refuting rumors by spreading accurate information about COVID-19 through television, the internet, newspapers, and other channels (Table 3).

Table 3. Chi-square test results of answers to the question "Eating a lot of garlic could prevent COVID-19."

Sources of COVID-19 information	Total participants, n	Participants who responded "yes," % (95% CI)	Participants who responded "no," % (95% CI)	χ^2 value	P value
Phone message or call	2206	44.10 (40.70-47.60)	46.50 (44.90-48.00)	1.48	.22
Suggestions from experts	3642	67.20 (64.00-70.40)	77.90 (76.50-79.10)	41.61	<.001
Video on social media	2056	46.00 (42.60-49.40)	42.30 (40.80-43.90)	3.68	.06
Opinions on social media	2230	45.90 (42.40-49.30)	46.70 (45.20-48.30)	0.20	.65
Newspapers	2771	50.90 (47.50-54.40)	59.30 (57.80-60.80)	19.27	<.001
Television	2871	55.70 (52.30-59.10)	60.80 (59.30-62.30)	7.20	.007
Friends	1238	27.80 (24.80-31.00)	25.50 (24.10-26.80)	1.94	.16
Patient's experience	405	10.50 (8.50-12.80)	8.00 (7.20-8.90)	5.27	.02
Family	1870	42.00 (38.70-45.50)	38.50 (36.90-40.00)	3.61	.06
Own COVID-19 experience	358	9.00 (7.20-11.10)	7.20 (6.40-8.00)	3.37	.07

Groups That Are Relatively Vulnerable to COVID-19 Misinformation

The results showed that residents from Central China (versus Eastern China, β=–0.342) and those who were divorced or widowed but not remarried (versus unmarried, β=–0.966) were significantly associated with a lower level of basic COVID-19 preventive perception (Table 4). In contrast, those aged 41-60 years (versus <20 years, β=0.9261), staff members in a government or public institution (versus seeking employment, β=0.821), education of >16 years (versus <6 years, β=1.294), and getting information about COVID-19 on television (β=0.683) were associated with higher scores (Table 4).

On the issue of identifying misleading information, students (versus those seeking employment, β=0.791), those with an education of >16 years (versus <6 years, β=0.774), those with one suspected symptom (versus those without suspected symptoms, β=0.494, P<.001), and those who take expert advice (β=0.322) were significantly associated with a higher ability to identify misleading information (Table 4). On the contrary, participants with >2 other diseases (versus those without other diseases, β=–0.601) and those with a tendency to get information from videos on social media (β=–0.589) were more likely to be indistinguishable from correct or incorrect information about COVID-19 (Table 4).

Table 4. Perception level of COVID-19 by demographic variable and knowledge source.

Variables	Basic COVID-19 preventive perception		Identified the misleading information	
	B	OR[a] (95% CI)	B	OR (95% CI)
Areas (reference: Eastern China)				
Central China	–0.342[b]	0.71 (0.51-1.00)	0.083	1.09 (0.89-1.32)
Western China	–0.200	0.82 (0.56-1.20)	0.071	1.07 (0.86-1.34)
Living in a rural area (reference: urban area)	–0.228	0.80 (0.59-1.09)	0.240[b]	1.27 (1.05-1.54)
Living alone (reference: living with others)	–0.373	0.69 (0.45-1.06)	0.202	1.22 (0.90-1.67)
Age, years (reference: ≤20)				
21-40	0.651[b]	1.92 (1.06-3.48)	0.121	1.13 (0.76-1.67)
41-60	0.926[c]	2.53 (1.25-5.12)	–0.009	0.99 (0.63-1.56)
>60	0.203	1.23 (0.55-2.74)	–0.041	0.96 (0.56-1.63)
Female (reference: male)	0.214	1.24 (0.95-1.61)	–0.174[b]	0.84 (0.72-0.99)
Marital status (reference: unmarried)				
Married or remarried	–0.574[b]	0.56 (0.32-0.98)	0.001	1.00 (0.72-1.39)
Divorced or widowed but not remarried	–0.966[c]	0.38 (0.18-0.79)	–0.119	0.89 (0.55-1.42)
Occupation (reference: seeking employment)				
Not working (not able to work)	0.205	1.23 (0.63-2.40)	0.090	1.09 (0.69-1.73)
Self-employed	0.260	1.30 (0.72-2.32)	0.050	1.05 (0.72-1.52)
Staff member in a government or public institution	0.821[b]	2.27 (1.11-4.65)	–0.061	0.94 (0.64-1.39)
Farmer, fisherman, or herdsman	0.138	1.15 (0.61-2.16)	–0.160	0.85 (0.56-1.31)
Retired	0.691	2.00 (0.99-4.05)	0.127	1.14 (0.73-1.76)
Student	0.639	1.90 (0.92-3.90)	0.79[c]	2.21 (1.41-3.46)
Staff member in a big company	0.365	1.44 (0.67-3.12)	0.118	1.13 (0.71-1.78)
Staff member in a small or medium company	0.604	1.83 (0.92-3.63)	0.387	1.47 (0.98-2.22)
Other	0.640	1.90 (0.98-3.68)	0.402	1.50 (0.99-2.26)
Household income in 2019, RMB (reference: <100,000 RMB)				
100,000-200,000	0.397[b]	1.49 (1.08-2.05)	0.274[b]	1.31 (1.09-1.59)
200,000-300,000	0.228	1.26 (0.77-2.05)	0.247	1.28 (0.97-1.69)
300,000-400,000	–0.21	0.81 (0.41-1.62)	0.466	1.59 (0.99-2.57)
>400,000	–0.082	0.92 (0.46-1.84)	0.208	1.23 (0.81-1.88)
Education, years (reference: ≤6 years)				
7-9	0.326	1.39 (0.91-2.10)	–0.038	0.96 (0.73-1.26)
10-12	0.208	1.23 (0.79-1.92)	–0.146	0.86 (0.65-1.15)
13-16	0.38	1.46 (0.89-2.40)	0.117	1.12 (0.82-1.54)
>16	1.294[b]	3.65 (1.17-11.34)	0.774[c]	2.17 (1.21-3.89)
Number of other diseases (reference: 0)				
1	0.625[c]	1.87 (1.21-2.88)	–0.040	0.96 (0.76-1.21)
2	0.145	1.16 (0.67-2.00)	–0.239	0.79 (0.57-1.09)
>2	0.079	1.08 (0.64-1.84)	–0.601[c]	0.55 (0.39-0.77)
Number of suspected symptoms (reference: 0)				

Variables	Basic COVID-19 preventive perception		Identified the misleading information	
	B	OR[a] (95% CI)	B	OR (95% CI)
1	–0.097	0.91 (0.55-1.51)	0.494[c]	1.64 (1.17-2.30)
2	0.486	1.63 (0.77-3.45)	0.288	1.33 (0.91-1.96)
>2	–0.294	0.75 (0.43-1.30)	–0.057	0.95 (0.66-1.35)
Number of COVID-19 information sources (reference: 0)				
1	0.750[b]	2.12 (1.14-3.93)	0.443[b]	1.56 (1.07-2.27)
2	0.396	1.49 (0.86-2.56)	0.219	1.25 (0.87-1.78)
3	0.695[b]	2.01 (1.10-3.65)	0.294	1.34 (0.92-1.95)
4	0.911[c]	2.49 (1.38-4.50)	0.434[b]	1.54 (1.05-2.26)
5	0.795[c]	2.25 (1.19-4.14)	0.911[c]	2.49 (1.59-3.90)
6	1.074[c]	2.93 (1.47-5.81)	0.764[c]	2.15 (1.36-3.40)
7	0.993[c]	2.70 (1.30-5.59)	0.433	1.54 (0.94-2.54)
8	0.804[b]	2.24 (1.07-4.68)	0.297	1.35 (0.79-2.30)
9	1.449[c]	4.26 (1.41-12.88)	0.662	1.94 (0.95-3.95)
Knowledge source				
Phone message or call	0.346	1.41 (0.94-2.13)	0.037	1.04 (0.82-1.31)
Suggestions from experts	0.28	1.32 (0.84-2.09)	0.322[b]	1.38 (1.03-1.84)
Video on social media	–0.404	0.67 (0.40-1.11)	–0.589[c]	0.55 (0.41-0.74)
Opinions on social media	–0.252	0.78 (0.47-1.30)	–0.038	0.96 (0.72-1.29)
Newspapers	-0.156	0.86 (0.54-1.36)	0.192	1.21 (0.93-1.58)
Television	0.683[c]	1.98 (1.31-2.99)	–0.056	0.95 (0.74-1.21)
Friends	–0.169	0.85 (0.54-1.33)	0.095	1.10 (0.84-1.43)
Patient's experience	–0.45	0.64 (0.34-1.19)	–0.248	0.78 (0.53-1.16)
Family	–0.217	0.81 (0.55-1.18)	–0.073	0.93 (0.74-1.16)
Own COVID-19 experience	–0.522	0.59 (0.32-1.12)	0.165	1.18 (0.77-1.82)
Constant	1.108	3.03	0.874	2.40

[a]OR: odds ratio.

[b]$P<.05$.

[c]$P<.001$.

Interestingly, the higher the number of information sources used by a participant, the higher their preventive perception and ability to identify misleading information. People who obtained information from >3 channels showed a significant improvement in their knowledge level compared to those who obtained information from only 1 or 2 channels (n=4, OR 2.49, 95% CI 1.38-4.50; n=9, OR 4.26, 95% CI 1.41-12.88). Videos on social media were the main culprit for the spreading of rumors, while expert advice helped to refute rumors, which is consistent with the results of the chi-square test above. The results of the stratified analysis showed that the population aged >60 years (OR 1.52, 95% CI 1.10-2.11), those with a lower- or middle-income level (OR 1.36, 95% CI 1.00-1.83), those who were not working and not able to work (OR 1.83, 95% CI 1.04-3.21), those with a household income <100,000 RMB (<US $14,954; OR 1.34, 95% CI 1.08-1.67), and those with >2 suspected symptoms (OR 2.95, 95% CI 1.50-5.80) were more likely to be misled by videos on social media (Figure 2). Although expert advice was the most useful tool to dispel rumors, its impact on vulnerable groups such as residents aged >60 years (OR 1.14, 95% CI 0.79-1.64), those who were divorced or widowed but not remarried (OR 0.54, 95% CI 0.25-1.16), those without work (OR 0.91, 95% CI 0.49-1.67), and those with suspected symptoms (OR 0.93, 95% CI 0.41-2.09) was limited (Table 5).

Figure 2. Vulnerable populations and their risk of being misled by incorrect information in videos on social media. OR: odds ratio.
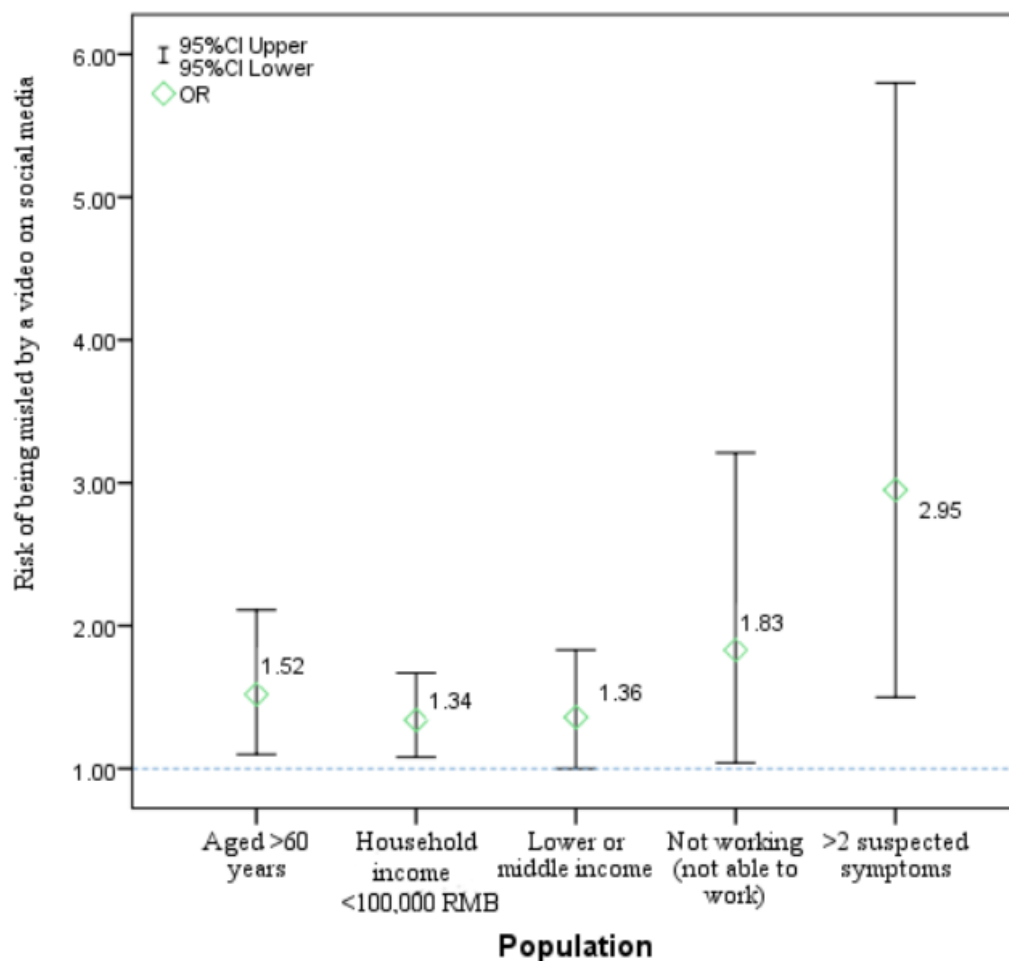
Table 5. Hierarchical analysis of the expert suggestion error correction effect among subgroups.

Characteristics	Participants who did not access expert advice	Participants who accessed expert advice	Odds ratio (95% CI)
	Participants who discerned rumors/total number of participants (%)	Participants who did not discern rumors/total number of participants (%)	
Age (years)			
≤20	115/135 (85.2)	419/464 (90.3)	1.62 (0.92-2.85)
21-40	277/358 (77.4)	1275/1416 (90.0)	2.64 (1.95-3.58)
41-60	331/441 (75.1)	941/1160 (81.1)	1.43 (1.10-1.85)
>60	158/212 (74.5)	463/602 (76.9)	1.14 (0.79-1.64)
Gender			
Male	450/570 (78.9)	1433/1678 (85.4)	1.56 (1.22-1.99)
Female	431/576 (74.8)	1665/1964 (84.8)	1.59 (1.36-1.86)
Marital status			
Unmarried	288/355 (81.1)	1248/1370 (91.1)	2.38 (1.72-3.29)
Married or remarried	546/734 (74.4)	1739/2117 (82.1)	1.58 (1.30-1.93)
Divorced or widowed but not remarried	47/57 (82.5)	111/155 (71.6)	0.54 (0.25-1.16)
Occupation			
Seeking employment	52/71 (73.2)	186/229 (81.2)	1.58 (0.85-2.94)
Not working (not able to work)	65/84 (77.4)	143/189 (75.7)	0.91 (0.49-1.67)
Self-employed shop owner or entrepreneur	128/172 (74.4)	323/397 (81.3)	1.50 (0.98-2.30)
Staff member in a government or public institution	107/141 (75.9)	398/474 (84.0)	1.66 (1.05-2.63)
Famer, fisherman, or herdsman	65/88 (73.9)	173/233 (74.2)	1.02 (0.58-1.78)
Retired	86/117 (73.5)	301/382 (78.8)	1.34 (0.83-2.16)
Student	184/220 (83.6)	869/935 (92.9)	2.58 (1.67-3.98)
Staff member in a big company	55/74 (74.3)	176/202 (87.1)	2.34 (1.20-4.55)
Staff member in a small or medium company	66/83 (79.5)	302/343 (88.0)	1.90 (1.02-3.54)
Other	73/96 (76.0)	227/258 (88.0)	2.31 (1.27-4.21)
Education (years)			
≤6	160/207 (77.3)	378/491 (77.0)	0.98 (0.67-1.45)
7-9	169/228 (74.1)	474/581 (81.6)	1.55 (1.08-2.22)
10-12	174/234 (74.4)	513/631 (81.3)	1.50 (1.05-2.14)
13-16	340/434 (78.3)	1519/1711 (88.8)	2.19 (1.66-2.88)
>16	38/43 (88.4)	214/228 (93.9)	2.01 (0.69-5.91)
Areas			
Eastern China	260/342 (76.0)	830/975 (85.1)	1.81 (1.33-2.45)
Central China	387/502 (77.1)	1433/1689 (84.8)	1.66 (1.30-2.13)
Western China	234/302 (77.5)	835/978 (85.4)	1.70 (1.23-2.34)
Type of area			
Urban	546/717 (76.2)	2000/2348 (85.2)	1.80 (1.47-2.21)
Rural	335/429 (78.1)	1098/1294 (84.9)	1.57 (1.19-2.07)
Household composition			
Living with others	782/1023 (76.4)	2843/3347 (84.9)	1.74 (1.46-2.07)
Living alone	99/123 (80.5)	255/295 (86.4)	1.55 (0.89-2.70)

XSL·FO
RenderX

Characteristics	Participants who did not access expert advice	Participants who accessed expert advice	Odds ratio (95% CI)
	Participants who discerned rumors/total number of participants (%)	Participants who did not discern rumors/total number of participants (%)	
Household income in 2019, RMB (US $)			
<100,000 (14,954)	406/524 (77.5)	1273/1550 (82.1)	1.34 (1.05-1.70)
100,000-200,000 (14,954-29,909)	323/417 (77.5)	1142/1318 (86.6)	1.89 (1.43-2.50)
200,000-300,000 (29,909-44,864)	100/132 (75.8)	390/447 (87.2)	2.19 (1.35-3.56)
300,000-400,000 (44,864-59,819)	17/26 (65.4)	153/167 (91.6)	5.79 (2.18-15.35)
>400,000 (>59,819)	35/47 (74.5)	140/160 (87.5)	2.40 (1.07-5.37)
Relative self-reported individual income			
Low (0%-20%)	287/373 (76.9)	952/1132 (84.1)	1.59 (1.19-2.12)
20%-40%	216/285 (75.8)	711/856 (83.1)	1.57 (1.13-2.17)
Average (40%-60%)	333/423 (78.7)	1214/1393 (87.2)	1.83 (1.38-2.43)
60%-80%	41/59 (69.5)	189/226 (83.6)	2.24 (1.16-4.33)
High (80%-100%)	4/6 (66.7)	32/35 (91.4)	5.33 (0.67-42.23)
Number of suspected symptoms			
0	745/981 (75.9)	2485/2933 (84.7)	1.76 (1.47-2.10)
1	67/75 (89.3)	272/307 (88.6)	0.93 (0.41-2.09)
2	39/48 (81.3)	178/204 (87.3)	1.58 (0.69-3.64)
>2	30/42 (71.4)	163/198 (82.3)	1.86 (0.87-3.99)
Number of other diseases			
0	662/843 (78.5)	2286/2626 (87.1)	1.84 (1.51-2.25)
1	137/172 (79.7)	487/597 (81.6)	1.13 (0.74-1.73)
2	44/64 (68.8)	187/232 (80.6)	1.89 (1.02-3.51)
>2	38/67 (56.7)	138/187 (73.8)	2.15 (1.20-3.85)

Discussion

Principal Findings

Overall, this study indicated a high level of awareness of COVID-19 among Chinese residents; this seems to be primarily related to education and the information sources used. More than half of the study sample had a bachelor's degree or higher. Our sample had a large number of students (24.1%) and staff in government or public institutions (12.8%), who have a high information acquisition ability, which also contributed to the high accuracy observed in this research. These results are similar to the results of an earlier study in China, in which participants of higher socioeconomic status were more knowledgeable and better able to take appropriate measures to prevent the spread of COVID-19 [22].

The ability to seek information from multiple channels due to the development of the information age had a significant impact. The higher the number of information sources, the more opportunity there is for individuals to acquire knowledge and consolidate existing knowledge; it is also likelier that individuals will be able to effectively discern misinformation from facts. Due to the seriousness of the epidemic, people actively sought

knowledge regarding the virus from various information sources, such as CCTV (China Central Television), the official website of the National Health Commission of China, and the official WeChat account of the Wuhan Health Commission [22]. For example, the "news 1+1" column takes the form of a dialogue between a host and an expert to answer the questions about the epidemic that people are most concerned about every day; it played a large role in the dissemination of knowledge and the refutation of rumors. Accurate information provided by trusted clinicians and scientists can help mitigate the spread of misinformation that is damaging to public health. As experts' suggestions played a role in correcting the "eating a lot of garlic could prevent COVID-19" rumor in our sample, health communication specialists may be able to directly counter prominent false narratives while promoting reliable sources of health information [26].

With the development of information technology, social media has played an important role in spreading information, including during and about this outbreak (46.6% and 42.9% of participants got information from opinions and videos on social media, respectively). Although these platforms provide easier and accessible ways of getting or generating information, they can

also be a source of misinformation [27]. Fake news on Weibo, WeChat, TikTok, and other short video platforms regarding potential drugs for COVID-19 (including ShuangHuangLian, garlic, and radix) resulted in unnecessary confusion and a shortage of drugs for patients who need them. Uncertainty breeds rumors and confusion, and social media platforms offer a fertile space for misinformation to be generated and disseminated [28]. This study highlighted the need for public health bodies to continue social media campaigns to minimize the circulation of inaccurate information about COVID-19 [23,29].

Among our participants, health perception and the ability to detect rumors was significantly lower among some vulnerable populations, including older participants and those with lower educational and income levels. These people may find it difficult to distinguish truth from misinformation in the news [30]. Meanwhile, they may be influenced by the traditional thinking pattern of "prefer to believe what they have rather than believe what they have not," which makes them more likely to be misled by rumors and to take preventive measures that do not have any effect. To make matters worse, although we found that listening to expert advice resulted in a higher ability to identify misleading information, expert advice plays a limited role in these vulnerable populations. Therefore, we conclude that health education interventions would be more effective if they focused on the targeted demographic groups [31]; for example, COVID-19 knowledge may be greatly increased if health education programs are specifically designed for older people and persons with a low level of education.

In this study, misleading questions were used to determine the main communication channels of misinformation and the groups that were most misled, which is more targeted than previous studies on knowledge level. However, this study still has some limitations. Rapid online surveys are a promising method to assess and track knowledge and perceptions in the midst of rapidly evolving infectious disease outbreaks [23]. An advantage of this study is that it used a rapid online survey to collect a large number of samples. However, a large proportion of participants were well-educated and engaged in mental work, resulting in a certain bias in the results. For this study, we recruited and trained the responsible investigators located in the sampled cities, who selected the sample by directly sending the questionnaire to families in their social network, which ensured we captured older adults in our sample. For those without the ability to use the internet, it is difficult to control the response process.

Conclusions

In general, our findings indicate that participants in the survey had a high level of health awareness of COVID-19 and were optimistic about success in the fight against the epidemic, which is important to limit the spread of the disease. Health information is spread through both traditional and new platforms, including television, newspapers, the internet, social media, and short video platforms, and the number of channels used to obtain information was positively associated with health perception. Among information sources, suggestions from experts are the most accurate source, while social media plays a large role in spreading rumors. We found that health knowledge was lower among older adults, those with less education, those who are unemployed or have a lower income, and those with underlying diseases. People in these groups were more likely to be misled by misinformation in videos on social media, while the error correction effect of experts was very limited in vulnerable populations. Although the government has taken major steps to limit the spread of the disease, more effort is needed to strengthen surveillance of social media, implement targeted support, and increase the influence of experts on vulnerable populations to reduce the spreading of rumors.

Conflicts of Interest

None declared.

References

1. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]

2. Zhou F. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3]

3. Lake M. What we know so far: COVID-19 current clinical knowledge and research. Clin Med (Lond) 2020 Mar;20(2):124-127 [FREE Full text] [doi: 10.7861/clinmed.2019-coron] [Medline: 32139372]

4. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 - Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/NEJMp2002125] [Medline: 32074416]

5. Anderson R, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020;395(10228):931-934. [doi: 10.1016/S0140-6736(20)30567-5]

6. Guo Y, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]

7. Del Rio C, Malani PN. COVID-19-New Insights on a Rapidly Changing Epidemic. JAMA 2020 Apr 14;323(14):1339-1340. [doi: 10.1001/jama.2020.3072] [Medline: 32108857]

8. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, HLH Across Speciality Collaboration, UK. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet 2020 Mar 28;395(10229):1033-1034 [FREE Full text] [doi: 10.1016/S0140-6736(20)30628-0] [Medline: 32192578]

9. Fisher D, Wilder-Smith A. The global community needs to swiftly ramp up the response to contain COVID-19. Lancet 2020 Apr 04;395(10230):1109-1110 [FREE Full text] [doi: 10.1016/S0140-6736(20)30679-6] [Medline: 32199470]

10. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-10-21]

11. World Health Organization. Coronavirus disease (COVID-) situation reports. 2019. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-10-21]

12. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, WHO StrategicTechnical Advisory Group for Infectious Hazards. COVID-19: towards controlling of a pandemic. Lancet 2020 Mar 28;395(10229):1015-1018 [FREE Full text] [doi: 10.1016/S0140-6736(20)30673-5] [Medline: 32197103]

13. Zhu F, Li YH, Guan XH, Hou LH, Wang WJ, Li J, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet 2020 Jun 13;395(10240):1845-1854 [FREE Full text] [doi: 10.1016/S0140-6736(20)31208-3] [Medline: 32450106]

14. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]

15. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]

16. Li J, You Z, Wang Q, Zhou ZJ, Qiu Y, Luo R, et al. The epidemic of 2019-novel-coronavirus (2019-nCoV) pneumonia and insights for emerging infectious diseases in the future. Microbes Infect 2020 Mar;22(2):80-85 [FREE Full text] [doi: 10.1016/j.micinf.2020.02.002] [Medline: 32087334]

17. Betsch C, Wieler LH, Habersaat K, COSMO group. Monitoring behavioural insights related to COVID-19. Lancet 2020 Apr 18;395(10232):1255-1256 [FREE Full text] [doi: 10.1016/S0140-6736(20)30729-7] [Medline: 32247323]

18. Tachfouti N, Slama K, Berraho M, Nejjari C. The impact of knowledge and attitudes on adherence to tuberculosis treatment: a case-control study in a Moroccan region. Pan Afr Med J 2012;12:52 [FREE Full text] [Medline: 22937192]

19. Abdelhafiz A, Mohammed Z, Ibrahim ME, Ziady HH, Alorabi M, Ayyad M, et al. Knowledge, Perceptions, and Attitude of Egyptians Towards the Novel Coronavirus Disease (COVID-19). J Community Health 2020 Oct;45(5):881-890 [FREE Full text] [doi: 10.1007/s10900-020-00827-7] [Medline: 32318986]

20. Hung L. The SARS epidemic in Hong Kong: what lessons have we learned? J R Soc Med 2003 Aug;96(8):374-378 [FREE Full text] [doi: 10.1258/jrsm.96.8.374] [Medline: 12893851]

21. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]

22. Zhong B, Luo W, Li HM, Zhang QQ, Liu XG, Li WT, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]

23. Geldsetzer P. Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]

24. Guideline for the diagnosis and treatment of COVID-19 (The Sixth Trial Version). Government of China. URL: http://www.gov.cn/zhengce/zhengceku/2020-02/19/content_5480948.htm [accessed 2020-10-21]

25. Covid-19 prevention and control programme (edition 6). Government of China. URL: http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtml [accessed 2020-10-21]

26. Vraga EK, Bode L. Using Expert Sources to Correct Health Misinformation in Social Media. Science Communication 2017;39(5):621-645. [doi: 10.1177/1075547017731776]

27. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]

28. Jacobsen K, Vraga EK. Improving communication about COVID-19 and emerging infectious diseases. Eur J Clin Invest 2020 May;50(5):e13225 [FREE Full text] [doi: 10.1111/eci.13225] [Medline: 32294248]

29.   Massaad E, Cherfan P. Social Media Data Analytics on Telehealth During the COVID-19 Pandemic. Cureus 2020 Apr 26;12(4):e7838 [FREE Full text] [doi: 10.7759/cureus.7838] [Medline: 32467813]
30.   Hightow-Weidman L, Muessig K, Claude K, Roberts J, Zlotorzynska M, Sanchez T. Maximizing Digital Interventions for Youth in the Midst of Covid-19: Lessons from the Adolescent Trials Network for HIV Interventions. AIDS Behav 2020 Aug;24(8):2239-2243 [FREE Full text] [doi: 10.1007/s10461-020-02870-w] [Medline: 32306214]
31.   van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244 [FREE Full text] [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]

Abbreviations

KAP: Knowledge-Attitude-Belief
WHO: World Health Organization

XSL•FO
RenderX

Review

Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature

Davide Golinelli[1*], MD, MPH; Erik Boetto[1*], MD; Gherardo Carullo[2], PhD, LLM; Andrea Giovanni Nuzzolese[3], PhD; Maria Paola Landini[4], MD; Maria Pia Fantini[1], MD

[1]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy

[2]Department of Italian and Supranational Public Law, University of Milan, Milan, Italy

[3]STLab, ISTC-CNR, Rome, Italy

[4]IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy

[*]these authors contributed equally

Corresponding Author:
Erik Boetto, MD
Department of Biomedical and Neuromotor Sciences
University of Bologna
via San Giacomo 12
Bologna, 40126
Italy
Phone: 39 0512094830
Email: erik.boetto@gmail.com

Abstract

Background: The COVID-19 pandemic is favoring digital transitions in many industries and in society as a whole. Health care organizations have responded to the first phase of the pandemic by rapidly adopting digital solutions and advanced technology tools.

Objective: The aim of this review is to describe the digital solutions that have been reported in the early scientific literature to mitigate the impact of COVID-19 on individuals and health systems.

Methods: We conducted a systematic review of early COVID-19–related literature (from January 1 to April 30, 2020) by searching MEDLINE and medRxiv with appropriate terms to find relevant literature on the use of digital technologies in response to the pandemic. We extracted study characteristics such as the paper title, journal, and publication date, and we categorized the retrieved papers by the type of technology and patient needs addressed. We built a scoring rubric by cross-classifying the patient needs with the type of technology. We also extracted information and classified each technology reported by the selected articles according to health care system target, grade of innovation, and scalability to other geographical areas.

Results: The search identified 269 articles, of which 124 full-text articles were assessed and included in the review after screening. Most of the selected articles addressed the use of digital technologies for diagnosis, surveillance, and prevention. We report that most of these digital solutions and innovative technologies have been proposed for the diagnosis of COVID-19. In particular, within the reviewed articles, we identified numerous suggestions on the use of artificial intelligence (AI)–powered tools for the diagnosis and screening of COVID-19. Digital technologies are also useful for prevention and surveillance measures, such as contact-tracing apps and monitoring of internet searches and social media usage. Fewer scientific contributions address the use of digital technologies for lifestyle empowerment or patient engagement.

Conclusions: In the field of diagnosis, digital solutions that integrate with traditional methods, such as AI-based diagnostic algorithms based both on imaging and clinical data, appear to be promising. For surveillance, digital apps have already proven their effectiveness; however, problems related to privacy and usability remain. For other patient needs, several solutions have been proposed, such as telemedicine or telehealth tools. These tools have long been available, but this historical moment may actually be favoring their definitive large-scale adoption. It is worth taking advantage of the impetus provided by the crisis; it is also important to keep track of the digital solutions currently being proposed to implement best practices and models of care in

XSL•FO

RenderX

future and to adopt at least some of the solutions proposed in the scientific literature, especially in national health systems, which have proved to be particularly resistant to the digital transition in recent years.

Introduction

Background

The COVID-19 pandemic, like all global crises in human history, is causing unprecedented health and economic disruptions in many countries. However, at the same time, this new situation is favoring the transition to digital solutions in many industries and in society as a whole. One example of this transition is education [1]; this entire sector, from primary schools to universities, has developed new strategies for teaching remotely, shifting from lectures in classrooms to live conferencing or web-based courses [2]. Similarly, health care organizations have responded to the COVID-19 pandemic through the rapid adoption of digital solutions and advanced technology tools. During a pandemic, digital technology can mitigate or even solve many challenges, thus improving health care delivery. Digital tools have been applied to address acute needs that have arisen as a direct or indirect consequence of the pandemic (eg, apps for patient tracing, remote triage emergency services). However, many of the solutions that have been developed and implemented during the emergency could be consolidated in the future, contributing to the definition and adoption of new digital models of care.

The list of new digital solutions is rapidly growing [3]. In addition to "video visits," these options include email and mobile phone apps as well as use of wearable devices, chatbots, artificial intelligence (AI)–powered diagnostic tools, voice-interface systems, and mobile sensors such as smart watches, oxygen monitors, or thermometers. A new category of service is the oversight of persons in home quarantine and large-scale population surveillance. Telemedicine and remote consultation have already proven to be effective at a time when access to health services for patients who do not have COVID-19 or for patients with nonacute COVID-19 is prevented, impeded, or postponed. In fact, according to Keesara et al [4], instead of using a model structured on the historically necessary model of in-person interactions between patients and their clinicians through a face-to-face model of care, current health care services and patient assistance can be guaranteed remotely through digital technologies.

Before the COVID-19 pandemic, it was expected that digital transformation in health care would be as disruptive as the transformations seen in other industries. However, as discussed by Hermann et al [5] and affirmed by Perakslis [6], "despite new technologies being constantly introduced, this change had yet to materialize." The spread of COVID-19 appears to have finally provided an ineludibly sound reason to fully embrace the digital transformation. Moreover, simulations show that many countries will probably face several waves of contagion, and new lockdowns will probably occur [7]. Therefore, it has become necessary to review the digital technologies that have been used during the emergency period and consider them for continued use over time or cyclically in the event of recurring outbreaks.

Goal of This Study

According to Hermann et al [5], digital technologies can be categorized based on the patient needs they address in health care: diagnosis, prevention, treatment, adherence, lifestyle, and patient engagement. We argue that it is necessary to understand which digital technologies have been adopted to face the COVID-19 crisis and whether and how they can still be useful after the emergency phase. To achieve this, it is crucial to cover as many aspects as possible of digital technology use in health care in response to the COVID-19 pandemic.

The aim of this study is therefore to describe the digital solutions that have been reported in the early scientific literature to mitigate the impact of COVID-19 on individuals and health systems.

Methods

Literature Search

We conducted a systematic review of the early scientific literature, following the Preferred Reporting Items for Systematic Reviews (PRISMA) approach [8], to include quantitative and qualitative studies using diverse designs to describe which digital solutions have been reported to respond and mitigate the effects of the COVID-19 pandemic. This review focuses on health research, which includes biomedical, epidemiological, clinical, public health, and health systems research.

The initial search was implemented on May 11, 2020, and was limited to the timespan from January 1 to April 30, 2020. The search query consisted of terms considered adequate by the authors to review the literature on the use of digital technologies in response to COVID-19. Therefore, we searched the MEDLINE database using the following search terms and database-appropriate syntax:

> *("COVID-19"[All Fields] OR "COVID-2019"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV-2"[All Fields] OR "2019nCoV"[All Fields] OR (("Wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND (2019/12[PDAT] OR 2020[PDAT]))) AND (digital[Title/Abstract] OR technology[Title/Abstract])]*

We also searched the COVID-19/SARS-CoV-2 section of medRxiv, a preprint server for health science papers that have yet to be peer-reviewed, for studies related to digital technologies, using the search string *COVID-19 digital technology* with the same timespan restriction applied to the MEDLINE search.

The search strategies and eligibility criteria used are provided in Multimedia Appendix 1.

Study Selection and Data Collection Process

We included articles for review if they were studies with original data or results referring to digital tools or interventions for COVID-19 and if they addressed the needs of patients or health care systems in the evaluation.

An article was excluded if it was not a study with original results; it did not focus on digital solutions for COVID-19; the full text was not available; or it was not written in English.

A two-stage screening process was used to assess the relevance of the identified studies. For the first level of screening, only the title and abstract were reviewed to preclude waste of resources in procuring articles that did not meet the minimum inclusion criteria. The titles and abstracts of the initially identified studies were checked by two independent investigators (DG and EB). For the second level of screening, all citations deemed relevant after the title and abstract screening were procured for subsequent review of the full-text article.

A spreadsheet in Excel (Microsoft Corporation) was developed to extract study characteristics such as the paper title, journal, publication date, type of technology, and patient needs addressed. In particular, we categorized the retrieved papers according to patient needs (diagnosis, prevention, treatment, adherence, lifestyle, and patient engagement). For the categorization of patient needs, we adapted the definition by Hermann et al [5], which reports the concept of "customer needs addressed" by the health care industry, to identify the patient health needs addressed by digital technology during the early phase of the COVID-19 pandemic.

The definition of patient needs is reported in Table 1. We added "surveillance" as an additional patient need to those identified by Hermann et al [5] given the importance of early identification and confinement of patients with COVID-19 to preserve population health, and a category of "other" to include any needs that were not considered in the previous categories.

Table 1. Definitions of the patient needs addressed by digital technologies.

Patient need	Definition
Diagnosis	The process of determining which disease or condition explains a person's symptoms and signs [9]
Prevention	Preventing the occurrence of a disease (eg, by reducing risk factors) or by halting a disease and averting resulting complications after its onset [10]
Adherence	The degree to which a patient correctly follows medical advice [11]
Treatment	The use of an agent, procedure, or regimen, such as a drug, surgery, or exercise, in an attempt to cure or mitigate a disease [12]
Lifestyle	Adoption and sustaining behaviors that can improve health and quality of life [13]
Patient engagement	Actively involving people in their health and health care [14]
Surveillance	The continuous, systematic collection, analysis, and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice [15]
Other	All patient needs, addressed by digital technology, which are not included in the previous categories

We built a scoring rubric by cross-classifying the patient needs addressed by the technology (or technologies) reported in each article with the type of technology itself. We relied on the report "Assessing the impact of digital transformation of Health Services" by the Expert Panel on Effective Ways of Investigating in Health (EXPH) of the European Commission [16] to classify the types of digital technologies (ie, AI, big data, chatbots, electronic health records [EHRs], mobile apps, robotics, sensors, telehealth, and telemedicine), integrating it with terms found within the analyzed articles when necessary (ie, blockchain, Internet of Things [IoT], internet search engines, social media, and mobile tracing).

We also extracted information and classified each technology reported by the selected articles according to health care system targets, grade of innovation, and scalability to other geographical areas. To do this, we also relied on the classifications and definitions reported by the EXPH (Table 2) [16].

Table 2. Classification of digital technologies and health services.

Classification category	Definition
Health care system targets	
Clients/patients	Members of the public who are potential or current users of health services, including caregivers.
Health care providers	Members of the health care workforce who deliver health services.
Health systems/resource managers	Systems and managers involved in the administration and oversight of public health systems. Interventions within this category reflect managerial functions related to supply chain management, health financing, and human resource management.
Data services	Crosscutting functionality to support a wide range of activities related to data collection, management, use, and exchange.
Grade of innovation	
Supporting	Digital services or technologies that can be used to support old or established technologies for all or some health care system targets. These technologies may support or facilitate the performance of existing technologies.
Complementing	Digital services or technologies that can be used in addition to old or established technologies for all or some health care system targets. These technologies may strengthen or enhance the performance of existing technologies.
Substituting	Digital services or technologies that may replace old or established technologies for all or some health care system targets.
Innovating	New digital services or technologies that may offer new possibilities that previously were not available for all or some health care system targets. These disruptive technologies may represent a new entry into the market.
Scalability to other geographical areas	
Not possible	Technologies strictly bonded to the context in which they were developed.
Local	Technologies whose scalability is limited to a local context (ie, regional or national context) for normative, legislative, ethical, or technical reasons.
Global	Technologies that do not present barriers to scalability that would prevent their possible global adoption.

Some of the analyzed articles described multiple technologies. For these articles, we reported all the health care system targets addressed by the proposed technologies. However, we found it impractical to assign different grades of innovation and scalability for each technology reported. Therefore, we chose to report only the highest grade of innovation or scalability assigned to the technologies within each article (eg, innovating>substituting>complementing>supporting).

Two of the authors (DG and EB) independently classified all identified articles in the predefined categories. Any disagreements were resolved through discussion and consensus between the two reviewers. If disagreement persisted, another reviewer (GC) was called as a tiebreaker.

Given the characteristics of this literature review, which aims to describe proposed digital solutions, and the nature and design of the included studies, assessments of the risk of bias and the study quality were not possible and therefore were not performed.

Results

Literature Search

The search identified 269 articles (174 from PubMed and 95 from medRxiv), of which 124 full-text articles were assessed and included in the review after screening (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews (PRISMA) literature review flowchart.
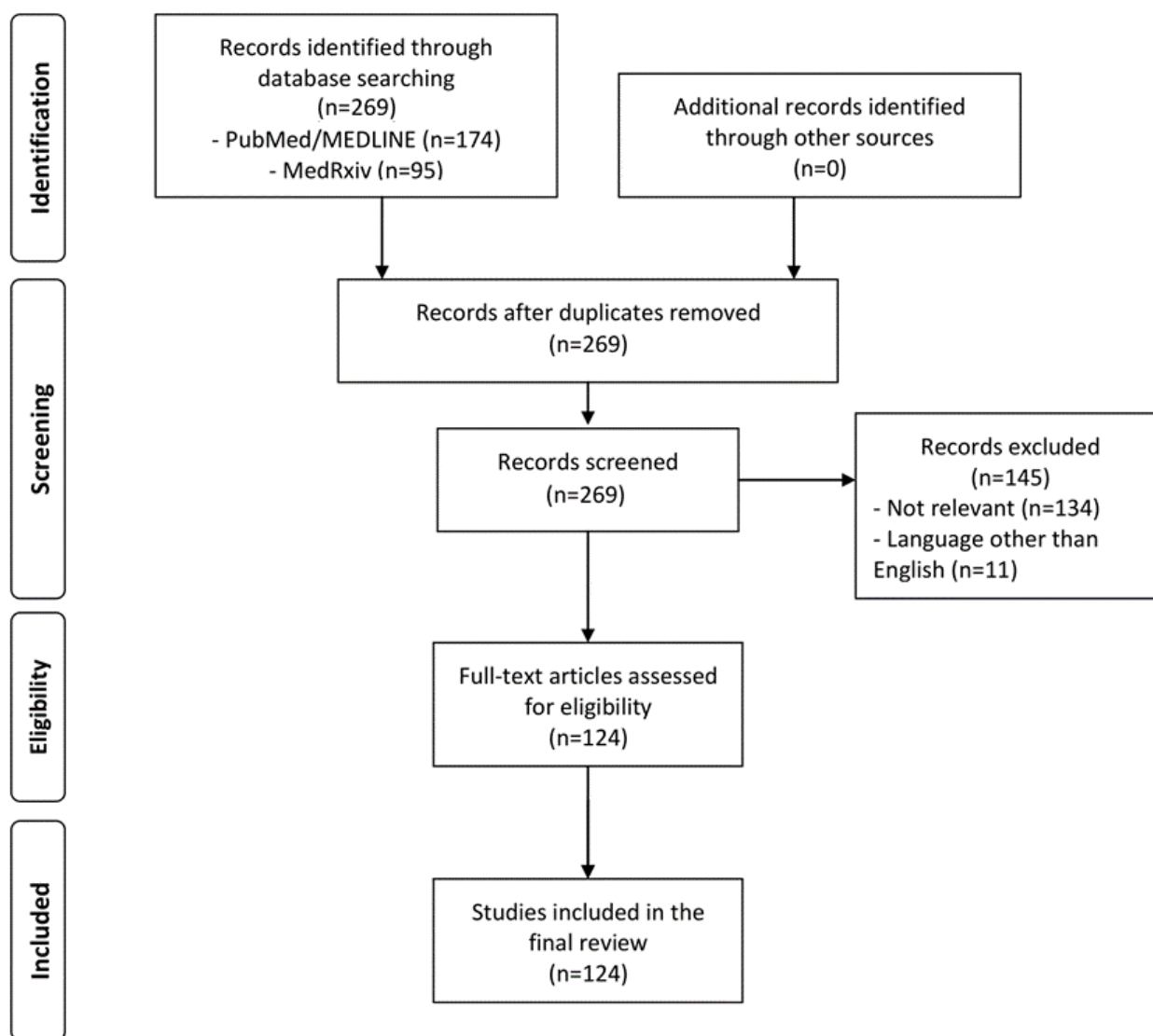
Study Selection and Data Collection Process

Out of the 124 selected articles, 65 (52.4%) addressed the use of digital technologies for diagnosis (Figure 2), 46 (37.1%) addressed surveillance, 46 (37.1%) addressed prevention, 38 (30.6%) addressed treatment, 15 (12.1%) addressed adherence, 12 (9.7%) addressed lifestyle, 11 (8.9%) addressed patient engagement, and 6 (4.8%) addressed other purposes. Considering the share of peer-reviewed articles, we found that for diagnosis, 39/65 articles (60%) were peer-reviewed; for surveillance, 29/46 (63%); for prevention, 30/46 (65%); for treatment, 33/38 (87%); for adherence, 15/15 (100%); for lifestyle, 11/12 (92%); for patient engagement, 11/11 (100%); and for other, 5 (83.3%).

In Table 3, we provide an extract of the characteristics of the included articles. An extended version of this table is provided in the supplementary material (Table S1, Multimedia Appendix 2).

XSL•FO
RenderX

Figure 2. Frequency of appearance of each patient need within the 124 selected articles and the share of peer-reviewed articles for each need. The total percentage is higher than 100 because some articles include technologies used to address more than one patient need.
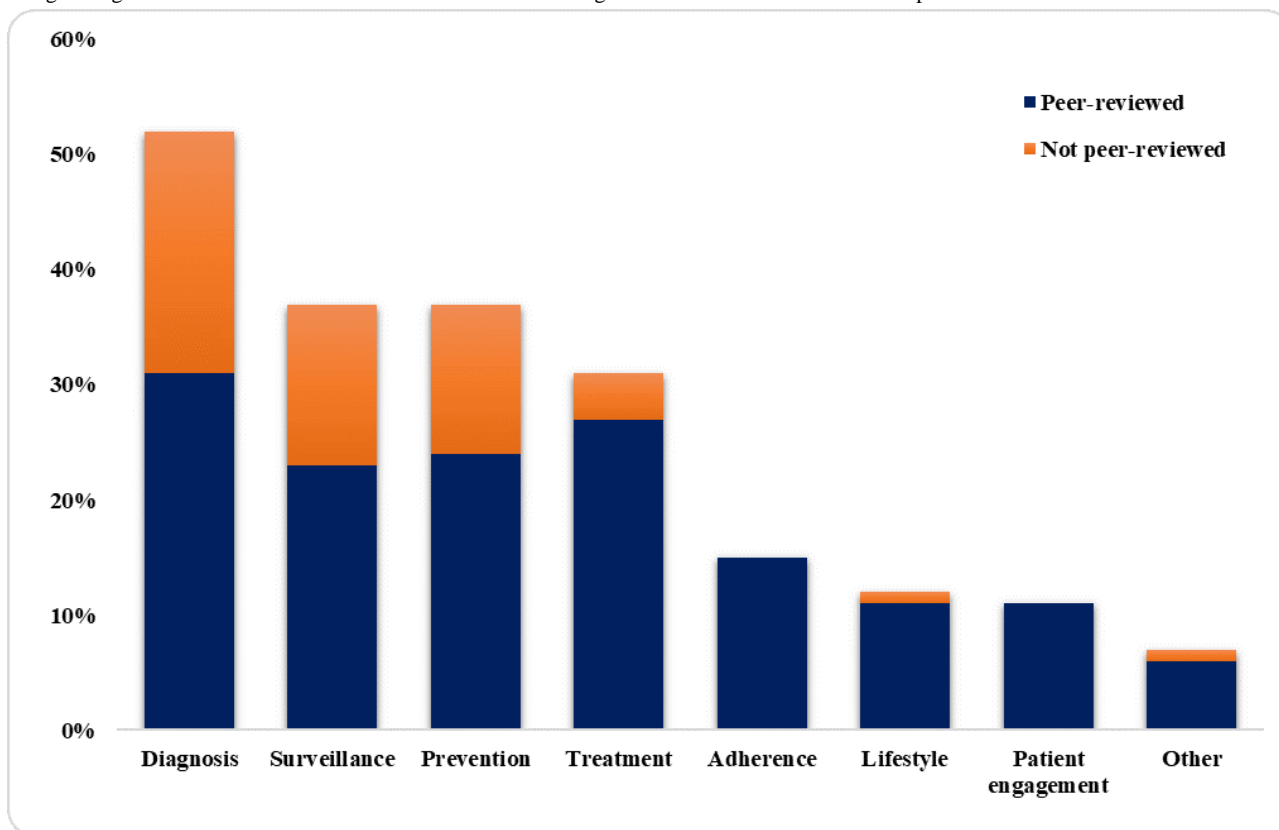
Table 3. Articles included in the literature review with the main characteristics of each analyzed paper.

ID	Reference	Health care system targets	Grade of innovation	Scalability
1	Zhai et al [17]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
2	Wang W et al [18]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
3	Wang et al [19]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
4	Yan et al [20]	Clients/patients, health care providers	Supporting	Global
5	Hou et al [21]	Clients/patients, health systems/resource managers, data services	Innovating	Local
6	Feng et al [22]	Health care providers	Complementing	Global
7	Jin et al [23]	Health care providers	Complementing	Global
8	Torous et al [24]	Clients/patients, health care providers, health systems/resource managers	Innovating	Global
9	Wang et al [25]	Health care providers	Complementing	Global
10	Zheng et al [26]	Health care providers	Complementing	Global
11	Galbiati et al [27]	Health care providers, health systems/resource managers	Supporting	Global
12	Bai et al [28]	Health care providers	Complementing	Global
13	Ienca et al [29]	Clients/patients, health systems/resource managers, data services	Innovating	Local
14	Ting et al [30]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
15	Hua et al [31]	Clients/patients, health systems/resource managers, data services	Innovating	Local
16	Fu et al [32]	Health care providers	Complementing	Global
17	Zhou et al [33]	Health care providers	Complementing	Global
18	Lin et al [34]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
19	Ferretti et al [35]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
20	Calton et al [36]	Clients/patients, health care providers	Innovating	Global
21	Lin et al [37]	Health care providers	Complementing	Global
22	Mashamba-Thompson et al [38]	Clients/patients, health systems/resource managers, data services	Innovating	Local
23	Salako et al [39]	Clients/patients, health care providers	Innovating	Local
24	Hernández et al [40]	Clients/patients, health systems/resource managers	Innovating	Local
25	Ohannessian et al [41]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
26	Shanlang et al [42]	Clients/patients, health systems/resource managers, data services	Innovating	Local
27	Turer et al [43]	Clients/patients, health care providers	Innovating	Global
28	Keesara et al [4]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
29	Her [44]	Clients/patients, health systems/resource managers, data services	Innovating	Local
30	Klum et al [45]	Clients/patients, health care providers	Supporting	Global
31	Calvo et al [46]	Clients/patients, health systems/resource managers, data services	Innovating	Local

XSL•FO
RenderX

ID	Reference	Health care system targets	Grade of innovation	Scalability
32	Dandekar et al [47]	Health systems/resource managers, data services	Innovating	Global
33	Drew et al [48]	Clients/patients, health systems/resource managers, data services	Innovating	Local
34	Segal et al [49]	Clients/patients, health systems/resource managers, data services	Innovating	Local
35	Hassanien et al [50]	Health care providers	Complementing	Global
36	Martin et al [51]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
37	Yasaka et al [52]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
38	Medford et al [53]	Clients/patients, health systems/resource managers, data services	Innovating	Local
39	Salg et al [54]	Health systems/resource managers, data services	Complementing	Local
40	Abhari et al [55]	Health systems/resource managers, data services	Innovating	Global
41	Jarynowski et al [56]	Clients/patients, health systems/resource managers, data services	Innovating	Local
42	Stommel et al [57]	Clients/patients, health care providers	Complementing	Global
43	Judson et al [58]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
44	Ćosić et al [59]	Clients/patients, health care providers	Innovating	Local
45	Grange et al [60]	Clients/patients, health care providers	Innovating	Global
46	Castiglioni et al [61]	Clinets/patients, health care providers	Complementing	Global
47	Serper et al [62]	Clients/patients, health care providers	Innovating	Global
48	Robbins et al [1]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
49	Crump [63]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
50	Punn et al [64]	Health systems/resource managers, data services	Innovating	Global
51	Myers et al [65]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
52	Noonan et al [66]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
53	Loebet et al [67]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
54	Price et al [68]	Health care providers	Innovating	Local
55	Vaishya et al [69]	Health care providers	Complementing	Global
56	Stubblefield et al [70]	Health care providers	Complementing	Global
57	Yuan et al [71]	Clients/patients, data services	Innovating	Local
58	Pollock et al [72]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
59	Mahmood et al [73]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
60	Gallotti et al [74]	Clients/patients, data services	Innovating	Local
61	Ren et al [75]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
62	Goldschmidt [76]	Clients/patients, health care providers	Innovating	Global
63	Al-karawi et al [77]	Health care providers	Complementing	Global
64	Kumar et al [78]	Health care providers	Complementing	Global

ID	Reference	Health care system targets	Grade of innovation	Scalability
65	Garg et al [79]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
66	Kuziemski et al [80]	Health systems/resource managers, data services	Innovating	Local
67	Jakhar et al [81]	Clients/patients, health care providers	Innovating	Local
68	Marasca et al [82]	Clients/patients, health care providers	Innovating	Local
69	Bulchandani et al [83]	Clients/patients, health systems/resource managers, data services	Innovating	Local
70	Green et al [84]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
71	Nagra et al [85]	Clients/patients, health care providers	Innovating	Global
72	O'Connor et al [86]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
73	Wittbold et al [87]	Clients/patients, health care providers	Innovating	Local
74	Hightow et al [88]	Clients/patients, health care providers	Innovating	Local
75	Bonavita et al [89]	Clients/patients, health care providers	Innovating	Local
76	Wosik et al [90]	Clients/patients, HO	Innovating	Global
77	Yan et al [91]	Health care providers, health systems/resource managers, data services	Supporting	Global
78	Lin et al [92]	Clients/patients, health care providers, health systems/resource managers, data services	Supporting	Not possible
79	Kummitha [93]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
80	Abbas et al [94]	Health care providers, data services	Complementing	Global
81	Ardabili et al [95]	Health systems/resource managers, data services	Supporting	Global
82	Zhang et al [96]	Clients/patients, health care providers,	Supporting	Global
83	Hart et al [97]	Clients/patients, health care providers	Innovating	Global
84	Parikh et al [98]	Clients/patients, health care providers	Innovating	Global
85	Rahman et al [99]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
86	Alwashmi Hart [100]	Clients/patients, health care providers, health systems/resource managers, data services	Complementing	Global
87	Sedov et al [101]	Health care providers, health systems/resource managers, data services	Complementing	Global
88	Mahapatra et al [102]	Health care providers; health systems/resource managers	Innovating	Global
89	Azizy et al [103]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
90	Madurai et al [104]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Global
91	Park et al [105]	Clients/patients, health care providers, health systems/resource managers, data services	Innovating	Local
92	Negrini et al [106]	Clients/patients, health care providers	Complementing	Global
93	Tanaka et al [107]	Clients/patients, health care providers	Complementing	Global
94	Randhawa et al [108]	Data services	Innovating	Global
95	Javaid et al [109]	Clients/patients, health care providers, health systems/resource managers, data services	Supporting	Global
96	Kyhlstedt et al [110]	Clients/patients, health care providers, health systems/resource managers, data services	Supporting	Global

ID	Reference	Health care system targets	Grade of innovation	Scalability
97	Barbosa et al [111]	Health systems/resource managers, data services	Innovating	Global
98	Blake et al [112]	Clients/patients, health systems/resource managers	Supporting	Global
99	Reeves et al [113]	Clients/patients, health care providers, health systems/resource managers, data services	Substituting	Global
100	Khan et al [114]	Clients/patients,	Substituting	Global
101	Whelan et al [115]	Clients/patients, health care providers, health systems/resource managers	Complementing	Global
102	Meinert et al [116]	Clients/patients, health care providers, data services	Innovating	Global
103	Ekong et al [117]	Clients/patients, health care providers, health systems/resource managers	Supporting	Local
104	Pérez Sust et al [118]	Clients/patients, health care providers, health systems/resource managers, data services	Complementing	Global
105	Kim et al [119]	Clients/patients, health care providers, health systems/resource managers, data services	Complementing	Global
106	Krukowski et al [120]	Clients/patients, health care providers, health systems/resource managers	Complementing	Global
107	Liu et al [121]	Health care providers, health systems/resource managers	Supporting	Global
108	Vaid et al [122]	Health care providers, health systems/resource managers	Complementing	Local
109	Lee [123]	Clients/patients, health care providers	Supporting	Global
110	Ramsetty et al [3]	Clients/patients, health care providers, health systems/resource managers	Complementing	Global
111	Tarek et al [124]	Health systems/resource managers	Complementing	Global
112	Awasthi et al [125]	Health care providers, health systems/resource managers	Supporting	Global
113	Khan et al [126]	Health care providers, health systems/resource managers	Supporting	Global
114	Husnayain et al [127]	Health care providers, health systems/resource managers	Supporting	Global
115	Weemaes et al [128]	Health care providers, health systems/resource managers, data services	Innovating	Local
116	Espinoza et al [129]	Clients/patients, health care providers, health systems/resource managers	Innovating	Global
117	Shweta et al [130]	Health care providers, health systems/resource managers, data services	Complementing	Global
118	Brat et al [131]	Health care providers, health systems/resource managers, data services	Supporting	Local
119	Hegde et al [132]	Clients/patients, health care providers	Innovating	Global
120	Tobias et al [133]	Clients/patients, health care providers, health systems/resource managers, data services	Supporting	Local
121	Compton et al [134]	Clients/patients, health care providers, health systems/resource managers	Complementing	Global
122	Smith et al [135]	Health care providers, health systems/resource managers	Complementing	Local
123	Kalteh et al [136]	Health care providers, health systems/resource managers, data services	Supporting	Global
124	Woo et al [137]	Clients/patients, health care providers	Complementing	Global

In Table 4, we cross-classify the use of technology and the type of technology. As an example, 24 articles describe the use of AI tools for the diagnosis of COVID-19, while 34 describe the use of telehealth or telemedicine for treatment purposes. All included articles and related analyses are reported in Multimedia Appendix 2 (Table S1).

Table 4. Cross-classification of the published studies by the type of technology and the patient needs addressed by the technology.

Technology	Diagnosis		Surveillance		Prevention		Treatment		Adherence		Lifestyle		Patient engagement		Other	
	n	Refs[a]	n	Refs	n	Refs	n	Refs	n	Refs	n	Refs	n	Refs	n	Refs
Artificial intelligence	24	[20,22,23,25,26, 28,30,32,33,38, 48,50,54,61,69, 70,77,78,94,100, 108,109,122,130]	12	[30,38,48,64,93, 95,100,104,109, 116,122,126]	11	[30,38,47,55,64, 95,100,109, 116,122,126]	2	[102,109]	1	[109]	1	[109]	1	[116]	4	[80, 104, 108, 125]
Big data analytics	6	[30,54,100,109, 118, 122]	11	[19,29,30,34,42, 93,100,109, 122,126, 136]	12	[19,30,34,42,55, 100,109,118, 122,124, 126,136]	2	[109,118]	2	[109,118]	3	[40, 109, 118]	0	N/A[b]	2	[111, 128]
Blockchain	2	[30,38]	2	[30,38]	2	[30,38]	0	N/A	0	N/A	0	N/A	0	N/A	0	N/A
Chatbot	3	[51,58,129]	0	N/A	1	[51]	0	N/A	0	N/A	0	N/A	0	N/A	0	N/A
Electronic health records	7	[1,60,91,113,118, 130,131]	5	[1,37,91,113,133]	4	[113,118,131,133]	5	[1,60,91, 113,118]	1	[118]	1	[118]	0	N/A	0	N/A
Internet of Things	3	[92,103,109]	5	[92,93,99, 104,109]	3	[92,103,109]	2	[103,109]	1	[109]	1	[109]	0	N/A	1	[104]
Internet search engines, social media	1	[1]	8	[1,18,21,31,42, 56,74,127]	3	[18,42,127]	1	[1]	0		1	[53]	0	N/A	1	[121]
Mobile app	8	[85,87,88,91,92, 96,100,110]	8	[18,91-93,96,100, 110,116]	9	[18,71,88,92,96, 100,110,115,116]	3	[85,88,91]	3	[88,112, 115]	2	[71, 112]	3	[112,115, 116]	0	N/A
Mobile tracing	6	[48,49,73,100, 110,118]	14	[31,35,44,46,48, 49,52,73,83,100, 105,110,117,119]	6	[44,73,83,100,110, 118]	2	[73,118]	1	[118]	1	[118]	0	N/A	0	N/A
Robotics, mechanical tools, drones, sensors, wearable devices	5	[45,101,109, 114,132]	3	[101,104,109]	3	[101,109,114]	2	[27,109]	3	[45,109, 114]	1	[109]	1	[114]	1	[104]
Telehealth, telemedicine	29	[1,17,39,57,59, 60,62,63,65-67, 72,73,75,76,81, 84-91,100,103, 107,118,120]	7	[1,73,75,89-91, 100]	18	[24,39,41, 43,71,73, 75,88-90, 100,103,115, 118,120,134, 135,137]	34	[1,3,17, 24,36,39, 57,59,60, 62,63,65 -68,72,73, 75,76,79, 84-86,88, 91,97,98, 103,106,107, 118,123,134, 135]	11	[24,82,88, 97,98,115, 118,120,134, 135,137]	8	[24, 65, 66, 71, 76, 79, 118, 137]	8	[59,76, 82,97, 115,120, 123,134]	0	N/A

[a]Refs: references.

[b]N/A: not applicable.

Characteristics of the Digital Health Technologies Favored by the COVID-19 Pandemic

Given the heterogeneity of the included technologies and solutions, we summarized the findings through a narrative synthesis. In fact, some patient needs share the same technology for health issues, which can be considered superimposable (eg, prevention and surveillance are often addressed by the same technology, such as mobile surveillance apps); this is also evident from Table 4, where many articles report technologies addressing more than one health need. Accordingly, in the following paragraphs, we describe the digital technologies and solutions starting from the patient needs they are designed to mitigate. We then discuss some of the retrieved articles that we deemed most interesting for each patient need.

Diagnosis

In the early scientific literature, digital solutions and innovative technologies were mainly proposed for the diagnosis of COVID-19. In particular, within the reviewed articles, we identified numerous suggestions of the use of AI-powered tools for the diagnosis and screening of SARS-CoV-2 or COVID-19, as reported in Table 4 [20,22,23,25,26,28,30,32,33, 38,48,50,54,61,69,70,77,78,94,100,108,109,122,130]. Most studies propose the adoption of AI tools based on the use of computed tomography (CT) data [23,28,33]. For example, Zhou et al [33] developed and validated an integrated deep learning framework on chest CT images for autodetection of novel coronavirus pneumonia (NCP), with particular focus on differentiating NCP from influenza pneumonia (IP), ensuring prompt implementation of isolation. Their AI model potentially provides an accurate early diagnostic tool for NCP. This type of diagnostic tool can be useful during the pandemic, especially when tests such as nucleic acid test kits are short in supply, which is a common problem during outbreaks. However, performing CT scans as a screening method presents significant limitations, considering the risks of both radiation exposure and operator or machine-type dependence [25].

In addition to these studies, many authors proposed AI-powered diagnostic tools for COVID-19 that are not based on CT scan data [22,51,58]. Feng et al [22] developed and validated a diagnosis aid model without CT images for early identification of suspected COVID-19 pneumonia on admission in adult fever patients and made the validated model available via a web-based triage calculator that must be supplied with clinical and serological data (eg, age, % monocytes, interleukin-6). Similarly, Martin et al [51] proposed a chatbot and a symptom-to-disease digital health assistant that can differentiate more than 20,000 diseases with an accuracy of more than 90%. The authors tested the accuracy of the digital health assistant to identify COVID-19 using a set of diverse clinical cases combined with case reports of COVID-19, and they reported that the digital health assistant can accurately distinguish COVID-19 in 96% of clinical cases.

A further innovative digital technology proposed to support the diagnosis of COVID-19 is blockchain (or distributed ledger) technology. In one study [38], the authors recommended low-cost blockchain and AI-coupled self-testing and tracking systems for COVID-19 and other emerging infectious diseases in low- and middle-income countries. They developed a low-cost blockchain app that requests a user's personal identifier before opening pretesting instructions. Following testing, the user uploads their results to the app, and the blockchain and AI systems enable the transfer to alert the outbreak surveillance. These types of solutions can also be of interest in high-income countries.

Another interesting digital tool proposed for the diagnosis and triage of patients is chatbots. Chatbots are applications that provide information through conversation-like interactions with users; they can be used for a broad range of purposes in health care (eg, patient triage, clinical decision support for providers, directing patients and staff to appropriate resources, and mental health applications). Chatbots can help evolve triage and screening processes in a scalable manner [129] and, as institutions become increasingly familiar with these tools, can be repurposed in the future for other public health emergencies as well as for more standard care uses.

Prevention and Surveillance

Our literature review suggests that digital technologies can be useful for COVID-19 diagnosis as well as for implementing prevention and surveillance measures.

Judson et al [58] deployed the Coronavirus Symptom Checker, a digital patient-facing self-triage and self-scheduling tool, to address the COVID-19 pandemic; it provides patients with 24-hour access to personalized recommendations, and it improves ambulatory surge capacity through self-triage, self-scheduling, and avoidance of unnecessary in-person care. The majority of patients who used this checker did not make any further contact with the health care system during subsequent days.

Another topic of paramount importance in the context of health care digitalization is epidemiological surveillance. Our review highlights that prevention and surveillance are often considered together in the scientific literature, given that "prevention of COVID-19" can be intended as "prevention of further spread," which is mainly achieved through surveillance. For the COVID-19 pandemic, surveillance definitely overlaps with prevention, as the risk of infection can be reduced by applying a successful surveillance plan and controlling the interactions between infected persons and the healthy population.

A study by Ferretti et al [35] analyzed the key parameters of the COVID-19 epidemic spread to estimate the contributions of different transmission routes and determine the requirements for successful case isolation and contact tracing. The viral spread is too rapid to be contained by manual contact tracing. The solution is the implementation of a contact-tracing app that creates a temporary record of proximity events between individuals and immediately alerts recent close contacts of diagnosed cases and prompts them to self-isolate. An important limitation of this type of tracing technology is that to achieve its goal, it must be used by a significant portion of the population.

An example of successful use of a mobile app for contact tracing is the app that the Chinese government implemented in Wuhan [31]. Quick response (QR) code–based screening was implemented in Hubei Province to monitor people's movement,

especially on public transportation. Using big data and mobile phones, three color codes were attributed to each citizen: green (safe), yellow (need to be cautious), and red (cannot enter). A similar tool was implemented in Taiwan [19]. In fact, through Taiwan citizens' household registration system and foreigners' entry cards, it was possible to track individuals at high risk of COVID-19 infection because of their recent history of travel to affected areas. If an individual was identified as high risk when in quarantine, they were monitored electronically through their mobile phone. Then, the Entry Quarantine System was launched: through the completion of a health declaration form (requiring the scan of a QR code that leads to a web-based form, either prior to departure from or upon arrival at a Taiwan airport), travelers could receive rapid immigration clearance.

Our literature review suggests that another meaningful way to control the spread of an epidemic is through monitoring and surveillance of internet searches and social media usage. Wang et al [18] used WeChat, a Chinese social media platform, to plot daily data on the frequencies of keywords related to SARS-CoV-2. The authors found that the frequencies of several keywords related to COVID-19 showed abnormal behavior during a period ahead of the outbreak in China, and they stated that social media can offer a new approach to early detection of disease outbreaks. Similarly, the Italian words for "cough" and "fever" were searched in Google Trends to find useful insights to predict the COVID-19 outbreak in Italy, showing a significant association with hospital admissions or deaths in the two following weeks [138]. These two papers show that tracking public health information from web-based search engines may have a role in the prediction of future COVID-19 outbreaks, complementarily to traditional public health surveillance systems.

Furthermore, a technology that can aid the automatic, decentralized, and remote collection of data for surveillance purposes is the IoT. In [99], an IoT-based smart disease surveillance system showed potential to control the pandemic. In fact, with most people using smartphones and wearable technologies and having internet access, this technology can help limit the spread of the pandemic through the collection and analysis of default gathered data.

Although its potential is irrefutable, the technology behind surveillance and contact tracing apps raises many concerns; as discussed by Calvo et al [46], the most obvious concern is "surveillance creep," which occurs when a surveillance tool developed for a precise goal (in the case of China and Taiwan, an app to monitor people's movement) remains in use after the crisis is solved. Privacy must be a primary concern for policy makers and a key challenge for designers and engineers who design digital tools for epidemic control. As already outlined in previous work by Carullo [139], in the European Union, applications to combat COVID-19 should avoid processing personal data whenever possible. The General Data Protection Regulation (GDPR) [140] dictates the principle of privacy by default, that is, "by default, only personal data which are necessary for each specific purpose of the processing are processed." In this regard, it should be reminded that according to the GDPR, data are "personal" only when and insofar they allow the identification of a natural person. Therefore, the processing of data, including clinical data, that cannot in any way identify a natural person does not involve personal data. Therefore, any privacy concerns are completely ruled out. To be compliant with this principle, a preferable approach is therefore to trace the spread of the virus, and therefore alert users, without collecting any personal data. A promising example in this direction was provided by Yasaka et al [52], who described an open-source proof-of-concept app for contact tracing that does not require registration or the divulgation of any private data, such as location. Instead, this tool uses an ingenious "checkpoint" system that allows users to create a peer-to-peer network of interactions and to determine if they have been exposed to any risk of infection; diagnosis of infection can be entered into the app, and the data are transferred to a central server but remain anonymous.

Although the aforementioned articles addressed surveillance and prevention in outpatients and the general population, an interesting point of view on inpatient surveillance comes from the study by Lin et al [37]. This paper describes a prospective active surveillance system with information technology (IT) services (ie, using a surveillance algorithm based on data from EHRs) to identify hospital inpatients whose pneumonia did not show marked improvement with antibiotic treatment and to alert the primary care medical teams on a daily basis. Similarly, EHR-based rapid screening processes, laboratory testing, clinical decision support, reporting tools, and patient-facing technology related to COVID-19 can be implemented using the EHR to build multiple COVID-19–specific tools to support outbreak management, including scripted triaging, electronic check-in, standard ordering and documentation, secure messaging, real-time data analytics, and telemedicine capabilities [113].

In the field of prevention, other important digital technologies proposed in the literature include telemedicine and telehealth. Telemedicine is not always applicable in emergencies, and many patients with COVID-19 may need to go to the hospital to receive higher level care. For this purpose, Turer et al [43] proposed using electronic personal protective equipment (ePPE) to protect staff (ie, prevent infection of health care workers) and conserve traditional PPE while providing rapid access to emergency care for low-risk patients during the COVID-19 pandemic. A very similar solution has been proposed by Wittbold et al [87]. They explored digital care delivery methods to reduce unnecessary exposure and conserve PPE through the deployment of iPads equipped with an app to evaluate and manage patients in a quaternary care academic and level one trauma center. Therefore, ePPE has been proved to be potentially applicable to settings such as emergency medical services, medical wards, and intensive care units.

Treatment and Adherence

Telemedicine and telehealth technologies are also used to increase patient adherence and for treatment purposes. Torous et al [24] describe the potential of digital health to increase access to and quality of mental health care by exploring the success of telehealth during the present crisis and how technologies such as apps can soon play larger roles. Telehealth is seen as a useful solution to deliver mental health care in general [141] and during social distancing and quarantine

periods. In addition, digital therapy programs can be offered through courses of evidence-based therapies or using augmented reality and virtual reality systems. As another example, Calton et al [36] deliver some useful advice on the implementation of telemedicine to deliver specialty palliative care in the homes of seriously ill patients and their families. The authors state that the digital divide must be taken into account. Patients require access to an internet connection and to a digital device suited for videoconferencing. For older people or those less familiar with technology, it may be necessary to identify a caregiver as a "technological liaison" for the patient. To create a successful telemedicine-based treatment environment, many critical factors must be considered, including workforce training, high-quality evidence, digital equity, and patient adherence.

Lifestyle and Patient Engagement

In the early literature responding to the COVID-19 pandemic, fewer scientific contributions addressed the use of digital technologies for lifestyle empowerment or patient engagement. This is probably due to the current phase of the pandemic, which has conditioned scientific research to focus primarily on aspects related to more acute patient needs. However, some articles can be found. For instance, Krukowski et al [120] addressed the issue of remote obesity management through telehealth methodologies such as electronic scales to remotely measure patients' weight and to maintain their engagement toward healthy lifestyles.

Discussion

Principal Results

Although SARS-CoV-2 is causing a pandemic worldwide, it is also favoring the rapid adoption of digital solutions and advanced technology tools in health care. On the one hand, physicians and health systems may need to track large populations of patients on a daily basis for surveillance purposes [4]. On the other hand, they may need rapid diagnostic tests for COVID-19 screening, to reduce the workload, and to enable patients to receive early diagnoses and timely treatments. This can also be achieved with the help of digital technologies, which were already available in different industries before the current crisis. These tools have now been quickly implemented in health care due to the pandemic [104].

In this systematic review of the early scientific literature in response to COVID-19, we describe numerous digital solutions and technologies addressing several patient and health care needs. The constantly updated scientific literature is a source of important ideas and suggestions for finding innovative solutions that guarantee patient care during and possibly after the COVID-19 crisis.

In the field of diagnosis, digital solutions that integrate with the traditional methods of clinical, molecular or serological diagnosis, such as AI-based diagnostic algorithms based both on imaging and clinical data, seem promising.

For surveillance, digital apps have already proven their effectiveness [142-144]; however, problems related to privacy and usability remain [145]. For other patient needs, several solutions have been proposed, such as telemedicine or telehealth

tools. These tools have long been available; however, this historical moment could actually favor their definitive large-scale adoption.

The fact that the digital technologies proposed in the analyzed scientific literature mainly address the fields of diagnosis, prevention, and surveillance probably reflects the emergency phase of the COVID-19 pandemic. As time passes, well-known digital tools could be proposed for different purposes and patient needs, such as adherence, lifestyle, and patient engagement, which are considered to be important determinants of patient health [146] despite the lower attention paid to them in the early scientific literature.

In addition to the patient needs addressed by digital technologies, our review sheds light on the most used digital technology tools. Given the early phase of the pandemic and its reflection on the articles included in this review, the technologies that have shown to be more easily and quickly implementable can be also considered as the most scalable. In fact, the speed with which these technologies have been deployed demonstrates their ease of adoption and manageability in many different contexts, despite their deployment during the course of a pandemic. Many of these solutions have demonstrated a technical, economic, regulatory and usability weight that is sufficiently low to allow their rapid and effective use, at least during the emergency phase. Among these solutions, we report AI tools for diagnosis, big data analytics and mobile tracing for surveillance and prevention, and telemedicine and telehealth, which have proved to be transversal tools for diagnosis, prevention, and treatment.

We advocate that many of the digital technologies that have been quickly implemented in this emergency phase can also be adopted in the following phases of the pandemic, as also stated by Fagherazzi et al [147]. However, this implementation is easier said than done. According to the article by Keesara et al [4], "Covid-19 and Health Care's Digital Revolution," in the context of the digital leap caused by the COVID-19 pandemic in the United States (and worldwide) [4], while private corporations and education institutions have made rapid transitions to remote work and videoconferencing, the health care system is still lagging behind in adopting digital solutions. This is mainly due to the fact that clinical workflows and economic incentives have been developed for a face-to-face model of care that, during this pandemic, contributes to the spread of the virus to uninfected patients who are seeking medical care. In addition to the history of health care policies, there are limiting factors to the implementation of tools such as telemedicine, including a legal framework that is not yet fully designed to regulate the use of innovative IT systems in health care, as well as an inadequate information and communications technology infrastructure and an obsolete reimbursement and payment structure.

Many countries are facing these regulatory issues: the challenges for digital health have become a global issue in the public health response to COVID-19 and future outbreaks. Digital tools such as telemedicine should indeed be integrated into international and national guidelines for public health preparedness, alongside the definition of national regulations and funding frameworks

in the context of public health emergencies. To switch to new digital models of care, increasing the digital expertise of health care professionals and educating the population are fundamental issues. Moreover, by implementing a data-sharing mechanism, digitally collected and stored data will be a precious tool for epidemiological surveillance that, as discussed earlier, is fundamental in controlling the epidemic spread. Lastly, to describe and assess the impact of digital tools during outbreaks, scientific evaluation frameworks should be defined [41].

Limitations

This literature review presents some limitations. First, the research was conducted in a period of epidemiological emergency. Thus, the number of daily publications is high, and it is difficult to keep up to date. As a result, we have been forced to select articles in a reduced time span, potentially missing some studies and including studies that have yet to be peer-reviewed. Secondly, due to the design of the review, the search could not be fully comprehensive, as it was conducted exclusively on the MEDLINE database and medRxiv to preserve

both time and resources; however, PubMed/MEDLINE is reported to be the primary database used by health science faculties [148], and medRxiv is one of the main repositories of COVID-19 research [149]. Finally, the articles and concepts included in this preliminary review certainly need to be integrated at the end of this international emergency phase.

Conclusions

The COVID-19 pandemic is favoring the implementation of digital solutions with unprecedented speed and impact. It is therefore recommended to keep track of the ideas and solutions being proposed today to implement best practices and models of care tomorrow and to be prepared for future national and international emergencies. It is worth taking advantage of the impetus provided by the crisis we are currently experiencing to implement at least some of the solutions proposed in the scientific literature, especially in national health systems, which in recent years have proved to be particularly resistant to the digital transition.

Multimedia Appendix 1
Eligibility criteria and search strategy.
[DOC File , 32 KB - jmir_v22i11e22280_app1.doc]

Multimedia Appendix 2
Table S1. Articles included in the literature review.
[XLSX File (Microsoft Excel File), 32 KB - jmir_v22i11e22280_app2.xlsx]

References

1. Robbins T, Hudson S, Ray P, Sankar S, Patel K, Randeva H, et al. COVID-19: A new digital dawn? Digital Health 2020;6:2055207620920083 [FREE Full text] [doi: 10.1177/2055207620920083] [Medline: 32313668]
2. Gewin V. Five tips for moving teaching online as COVID-19 takes hold. Nature 2020 Apr;580(7802):295-296. [doi: 10.1038/d41586-020-00896-7] [Medline: 32210377]
3. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
4. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
5. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study. J Med Internet Res 2018 Mar 27;20(3):e104 [FREE Full text] [doi: 10.2196/jmir.9498] [Medline: 29588274]
6. Perakslis ED. Strategies for delivering value from digital technology transformation. Nat Rev Drug Discov 2017 Feb;16(2):71-72. [doi: 10.1038/nrd.2016.265] [Medline: 28082744]

7. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]

8. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]

9. Langlois JP. Making a Diagnosis. In: Mengel MB, Holleman WL, Fields SA, editors. Fundamentals of Clinical Practice, 2nd Edition. Boston, MA: Springer; 2002:197-217.

10. Clarke EA. What is Preventive Medicine? Can Fam Physician 1974 Nov;20(11):65-68 [FREE Full text] [Medline: 20469128]

11. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.

12. Drexler M, Institute of Medicine. Prevention and Treatment. In: What You Need to Know About Infectious Disease. Washington, DC: National Academies Press; 2010.

13. Lianov L, Johnson M. Physician competencies for prescribing lifestyle medicine. JAMA 2010 Jul 14;304(2):202-203. [doi: 10.1001/jama.2010.903] [Medline: 20628134]

14. James J. Patient Engagement. Health Affairs. 2013 Feb 14. URL: https://www.healthaffairs.org/do/10.1377/hpb20130214.898775/full/ [accessed 2020-11-03]

15. Public health surveillance. World Health Organization. URL: https://www.who.int/immunization/monitoring_surveillance/burden/vpd/en/ [accessed 2020-11-03]

16. European Commission. Assessing the Impact of Digital Transformation of Health Services: Report of the Expert Panel on effective ways of investing in Health (EXPH). Luxembourg: Publications Office of the European Union; Nov 2019:ckz185.769.

17. Zhai Y, Wang Y, Zhang M, Gittell JH, Jiang S, Chen B, et al. From Isolation to Coordination: How Can Telemedicine Help Combat the COVID-19 Outbreak? medRxiv Preprint posted online on February 23, 2020. [doi: 10.37473/dac/10.1101/2020.02.20.20025957]

18. Wang W, Wang Y, Zhang X, Li Y, Jia X, Dang S. WeChat, a Chinese social media, may early detect the SARS-CoV-2 outbreak in 2019. medRxiv Preprint posted online on February 26, 2020. [doi: 10.1101/2020.02.24.20026682]

19. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]

20. Yan L, Zhang H, Xiao Y, Wang M, Sun C, Liang J, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv Preprint posted online on March 03, 2020. [doi: 10.1101/2020.02.27.20028027]

21. Hou Z, Du F, Jiang H, Zhou X, Lin L. Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China. medRxiv Preprint posted online on March 17, 2020. [doi: 10.1101/2020.03.14.20035956]

22. Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, et al. A Novel Triage Tool of Artificial Intelligence Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics. SSRN Journal Preprint posted online on March 19, 2020. [doi: 10.2139/ssrn.3551355]

23. Jin S, Wang B, Xu H, Wei L, Zhao W, Hou X, et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. medRxiv Preprint posted online on March 23, 2020. [doi: 10.1101/2020.03.19.20039354]

24. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]

25. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. medRxiv Preprint posted online on March 26, 2020. [doi: 10.1101/2020.03.24.20042317]

26. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label. medRxiv Preprint posted online on March 26, 2020. [doi: 10.1101/2020.03.12.20027185]

27. Galbiati C, Bonivento W, Caravati M, Razeti M, DeCecco S, Fiorillo G, et al. Mechanical Ventilator Milano (MVM): A Novel Mechanical Ventilator Designed for Mass Scale Production in Response to the COVID-19 Pandemics. medRxiv Preprint posted online on March 27, 2020. [doi: 10.1101/2020.03.24.20042234]

28. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, et al. Predicting COVID-19 Malignant Progression with AI Techniques. SSRN Journal Preprint posted online on March 31, 2020. [doi: 10.2139/ssrn.3557984]

29. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]

30. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]

31. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]

32. Fu M, Yi S, Zeng Y, Ye F, Li Y, Dong X, et al. Deep Learning-Based Recognizing COVID-19 and other Common Infectious Diseases of the Lung by Chest CT Scan Images. medRxiv Preprint posted online on March 30, 2020. [doi: 10.1101/2020.03.28.20046045]

33. Zhou M, Chen Y, Wang D, Xu Y, Yao W, Huang J, et al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. medRxiv Preprint posted online on March 30, 2020. [doi: 10.1101/2020.03.24.20043117]

34. Lin C, Braund WE, Auerbach J, Chou J, Teng J, Tu P, et al. Policy Decisions and Use of Information Technology to Fight COVID-19, Taiwan. Emerg Infect Dis 2020 Jul;26(7):1506-1512 [FREE Full text] [doi: 10.3201/eid2607.200574] [Medline: 32228808]

35. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]

36. Calton B, Abedini N, Fratkin M. Telemedicine in the Time of Coronavirus. J Pain Symptom Manage 2020 Jul;60(1):e12-e14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]

37. Lin C, Cheng C, Lu P, Shih D, Hung C, Lo H, et al. Active surveillance for suspected COVID-19 cases in inpatients with information technology. J Hosp Infect 2020 Jun;105(2):197-199. [doi: 10.1016/j.jhin.2020.03.027] [Medline: 32243950]

38. Mashamba-Thompson TP, Crayton ED. Blockchain and Artificial Intelligence Technology for Novel Coronavirus Disease-19 Self-Testing. Diagnostics (Basel) 2020 Apr 01;10(4):198 [FREE Full text] [doi: 10.3390/diagnostics10040198] [Medline: 32244841]

39. Salako O, Okunade K, Allsop M, Habeebu M, Toye M, Oluyede G, et al. Upheaval in cancer care during the COVID-19 outbreak. Ecancermedicalscience 2020;14:ed97 [FREE Full text] [doi: 10.3332/ecancer.2020.ed97] [Medline: 32269597]

40. Hernández-García I, Giménez-Júlvez T. Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]

41. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]

42. Lin S, Ma C, Lin R, Huang J, Xu R, Yuan A. Research on the Influence of Information Diffusion on the Transmission of the Novel Coronavirus (COVID-19). medRxiv Preprint posted online on April 02, 2020. [doi: 10.1101/2020.03.31.20048439]

43. Turer RW, Jones I, Rosenbloom ST, Slovis C, Ward MJ. Electronic personal protective equipment: A strategy to protect emergency department providers in the age of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):967-971 [FREE Full text] [doi: 10.1093/jamia/ocaa048] [Medline: 32240303]

44. Her M. How Is COVID-19 Affecting South Korea? What Is Our Current Strategy? Disaster Med Public Health Prep 2020 Apr 03:1-3 [FREE Full text] [doi: 10.1017/dmp.2020.69] [Medline: 32241325]

45. Klum M, Urban M, Tigges T, Pielmus A, Feldheiser A, Schmitt T, et al. Wearable Cardiorespiratory Monitoring Employing a Multimodal Digital Patch Stethoscope: Estimation of ECG, PEP, LVETand Respiration Using a 55 mm Single-Lead ECG and Phonocardiogram. Sensors (Basel) 2020 Apr 04;20(7):2033 [FREE Full text] [doi: 10.3390/s20072033] [Medline: 32260436]

46. Calvo R, Deterding C, Ryan R. Health surveillance during covid-19 pandemic. BMJ 2020 Apr 06;369:m1373. [doi: 10.1136/bmj.m1373] [Medline: 32253180]

47. Dandekar R, Barbastathis G. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. medRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.04.03.20052084]

48. Drew D, Nguyen L, Steves C, Wolf J, Spector TD, Chan AT, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. medRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.04.02.20051334]

49. Segal E, Zhang F, Lin X, King G, Shalem O, Shilo S, CCC (Coronavirus Census Collective). Building an international consortium for tracking coronavirus health status. medRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.04.02.20051284]

50. Hassanien AS, Mahdy LS, Ezzat KS, Elmousalami HS, Ella HJ. Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine. medRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.03.30.20047787]

51. Martin A, Nateqi J, Gruarin S. An artificial intelligence based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. bioRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.03.25.008805]

52. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]

53. Medford R, Saleh S, Sumarsono A, Perl TM, Lehmann CU. An "Infodemic": Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak. medRxiv Preprint posted online on April 07, 2020. [doi: 10.1101/2020.04.03.20052936]

54. Salg G, Ganten M, Baumhauer M, Heussel C, Kleesiek J. A globally available COVID-19 - Template for clinical imaging studies. medRxiv Preprint posted online on April 07, 2020. [FREE Full text] [doi: 10.1101/2020.04.02.20048793]

55. Abhari R, Marini M, Chokani N. COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data. medRxiv Preprint posted online on April 07, 2020. [doi: 10.1101/2020.03.30.20047472]

56. Jarynowski A, Wojta-Kempa M, Belik V. Perception of emergent epidemic of COVID-2019 / SARS CoV-2 on the Polish Internet. medRxiv Preprint posted online on April 07, 2020. [doi: 10.1101/2020.03.29.20046789]

57. Stommel WJP, van Goor H, Stommel MWJ. The Impact of Video-Mediated Communication on Closed Wound Assessments in Postoperative Consultations: Conversation Analytical Study. J Med Internet Res 2020 May 05;22(5):e17791 [FREE Full text] [doi: 10.2196/17791] [Medline: 32310816]

58. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

59. Ćosić K, Popović S, Šarlija M, Kesedžić I. Impact of Human Disasters and COVID-19 Pandemic on Mental Health: Potential of Digital Psychiatry. Psychiatr Danub 2020;32(1):25-31 [FREE Full text] [doi: 10.24869/psyd.2020.25] [Medline: 32303026]

60. Grange ES, Neil EJ, Stoffel M, Singh AP, Tseng E, Resco-Summers K, et al. Responding to COVID-19: The UW Medicine Information Technology Services Experience. Appl Clin Inform 2020 Mar;11(2):265-275 [FREE Full text] [doi: 10.1055/s-0040-1709715] [Medline: 32268390]

61. Castiglioni I, Ippolito D, Interlenghi M, Monti CB, Salvatore C, Schiaffino S, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy, Italy. medRxiv Preprint posted online on April 10, 2020. [FREE Full text] [doi: 10.1101/2020.04.08.20040907]

62. Serper M, Cubell AW, Deleener ME, Casher TK, Rosenberg DJ, Whitebloom D, et al. Telemedicine in Liver Disease and Beyond: Can the COVID-19 Crisis Lead to Action? Hepatology 2020 Aug;72(2):723-728 [FREE Full text] [doi: 10.1002/hep.31276] [Medline: 32275784]

63. Crump WJ. Telemedicine: Has the Time Really Finally Arrived? J Rural Health 2020 Apr 11 [FREE Full text] [doi: 10.1111/jrh.12435] [Medline: 32277729]

64. Punn N, Sonbhadra S, Agarwal S. COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. medRxiv Preprint posted online on June 01, 2020. [doi: 10.1101/2020.04.08.20057679]

65. Myers US, Birks A, Grubaugh AL, Axon RN. Flattening the Curve by Getting Ahead of It: How the VA Healthcare System Is Leveraging Telehealth to Provide Continued Access to Care for Rural Veterans. J Rural Health 2020 Apr 13 [FREE Full text] [doi: 10.1111/jrh.12449] [Medline: 32282955]

66. Noonan D, Simmons LA. Navigating Nonessential Research Trials During COVID-19: The Push We Needed for Using Digital Technology to Increase Access for Rural Participants? J Rural Health 2020 Apr 13 [FREE Full text] [doi: 10.1111/jrh.12446] [Medline: 32282959]

67. Loeb AE, Rao SS, Ficke JR, Morris CD, Riley LH, Levin AS. Departmental Experience and Lessons Learned With Accelerated Introduction of Telemedicine During the COVID-19 Crisis. J Am Acad Orthop Surg 2020 Jun 01;28(11):e469-e476 [FREE Full text] [doi: 10.5435/JAAOS-D-20-00380] [Medline: 32301818]

68. Price A, Henke L, Maraghechi B, Kim T, Spraker MB, Hugo GD, et al. Implementation of a Novel Remote Physician SBRT Coverage Process during the Coronavirus Pandemic. medRxiv Preprint posted online on April 14, 2020. [doi: 10.1101/2020.04.09.20059857]

69. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):337-339 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.012] [Medline: 32305024]

70. Stubblefield J, Hervert M, Causey J, Qualls J, Dong W, Cai L, et al. Cardiac or Infectious? Transfer Learning with Chest X-Rays for ER Patient Classification. medRxiv Preprint posted online on April 15, 2020. [FREE Full text] [doi: 10.1101/2020.04.11.20062091]

71. Yuan EJ, Hsu C, Lee W, Chen T, Chou L, Hwang S. Where to buy face masks? Survey of applications using Taiwan's open data in the time of coronavirus disease 2019. J Chin Med Assoc 2020 Jun;83(6):557-560 [FREE Full text] [doi: 10.1097/JCMA.0000000000000325] [Medline: 32304508]

72. Pollock K, Setzen M, Svider PF. Embracing telemedicine into your otolaryngology practice amid the COVID-19 crisis: An invited commentary. Am J Otolaryngol 2020;41(3):102490 [FREE Full text] [doi: 10.1016/j.amjoto.2020.102490] [Medline: 32307192]

73. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]

74. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of ?infodemics? in response to COVID-19 epidemics. medRxiv Preprint posted online on April 16, 2020. [FREE Full text] [doi: 10.1101/2020.04.08.20057968]

75. Ren X, Zhai Y, Song X, Wang Z, Dou D, Li Y. The Application of Mobile Telehealth System to Facilitate Patient Information Presentation and Case Discussion. Telemed J E Health 2020 Jun;26(6):725-733. [doi: 10.1089/tmj.2020.0084] [Medline: 32298208]

76. Goldschmidt K. The COVID-19 Pandemic: Technology use to Support the Wellbeing of Children. J Pediatr Nurs 2020;53:88-90 [FREE Full text] [doi: 10.1016/j.pedn.2020.04.013] [Medline: 32317129]

77. Al-karawi D, Al-Zaidi S, Polus N, Jassim S. Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients. medRxiv Preprint posted online on April 17, 2020. [doi: 10.1101/2020.04.13.20063479]

78. Kumar R, Arora A, Bansal B, Sahayasheela VJ, Buckchash H, Imran J, et al. Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers. medRxiv Preprint posted online on April 17, 2020. [doi: 10.1101/2020.04.13.20063461]

79. Garg SK, Rodbard D, Hirsch IB, Forlenza GP. Managing New-Onset Type 1 Diabetes During the COVID-19 Pandemic: Challenges and Opportunities. Diabetes Technol Ther 2020 Jun;22(6):431-439. [doi: 10.1089/dia.2020.0161] [Medline: 32302499]

80. Kuziemski M, Misuraca G. AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecomm Policy 2020 Jul;44(6):101976 [FREE Full text] [doi: 10.1016/j.telpol.2020.101976] [Medline: 32313360]

81. Jakhar D, Kaur I, Kaul S. Screen mirroring, screen casting and screen sharing during COVID-19: what dermatologists should know. Clin Exp Dermatol 2020 Aug;45(6):750-751 [FREE Full text] [doi: 10.1111/ced.14247] [Medline: 32304574]

82. Marasca C, Ruggiero A, Fontanella G, Ferrillo M, Fabbrocini G, Villani A. Telemedicine and support groups could be used to improve adherence to treatment and health-related quality of life in patients affected by inflammatory skin conditions during the COVID-19 pandemic. Clin Exp Dermatol 2020 Aug;45(6):749 [FREE Full text] [doi: 10.1111/ced.14245] [Medline: 32304587]

83. Bulchandani V, Shivam S, Moudgalya S, Sondhi S. Digital Herd Immunity and COVID-19. medRxiv Preprint posted online on May 26, 2020. [doi: 10.1101/2020.04.15.20066720]

84. Green A, Ruchman S, Katz C, Singer E. Piloting Forensic Tele-Mental Health Evaluations of Asylum Seekers. medRxiv Preprint posted online on April 18, 2020. [doi: 10.1101/2020.04.15.20063677]

85. Nagra M, Vianya-Estopa M, Wolffsohn JS. Could telehealth help eye care practitioners adapt contact lens services during the COVID-19 pandemic? Cont Lens Anterior Eye 2020 Jun;43(3):204-207 [FREE Full text] [doi: 10.1016/j.clae.2020.04.002] [Medline: 32336578]

86. O'Connor CM, Anoushiravani AA, DiCaprio MR, Healy WL, Iorio R. Economic Recovery After the COVID-19 Pandemic: Resuming Elective Orthopedic Surgery and Total Joint Arthroplasty. J Arthroplasty 2020 Jul;35(7S):S32-S36 [FREE Full text] [doi: 10.1016/j.arth.2020.04.038] [Medline: 32345566]

87. Wittbold KA, Baugh JJ, Yun BJ, Raja AS, White BA. iPad deployment for virtual evaluation in the emergency department during the COVID-19 pandemic. Am J Emerg Med 2020 Apr 18:0` [FREE Full text] [doi: 10.1016/j.ajem.2020.04.025] [Medline: 32317204]

88. Hightow-Weidman L, Muessig K, Claude K, Roberts J, Zlotorzynska M, Sanchez T. Maximizing Digital Interventions for Youth in the Midst of Covid-19: Lessons from the Adolescent Trials Network for HIV Interventions. AIDS Behav 2020 Aug;24(8):2239-2243 [FREE Full text] [doi: 10.1007/s10461-020-02870-w] [Medline: 32306214]

89. Bonavita S, Tedeschi G, Atreja A, Lavorgna L. Digital triage for people with multiple sclerosis in the age of COVID-19 pandemic. Neurol Sci 2020 May;41(5):1007-1009 [FREE Full text] [doi: 10.1007/s10072-020-04391-9] [Medline: 32303856]

90. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]

91. Yan A, Zou Y, Mirchandani DA. How hospitals in mainland China responded to the outbreak of COVID-19 using information technology-enabled services: An analysis of hospital news webpages. J Am Med Inform Assoc 2020 Jul 01;27(7):991-999 [FREE Full text] [doi: 10.1093/jamia/ocaa064] [Medline: 32311036]

92. Lin B, Wu S. COVID-19 (Coronavirus Disease 2019): Opportunities and Challenges for Digital Health and the Internet of Medical Things in China. OMICS 2020 May;24(5):231-232. [doi: 10.1089/omi.2020.0047] [Medline: 32311300]

93. Kummitha RKR. Smart technologies for fighting pandemics: The techno- and human- driven approaches in controlling the virus transmission. Gov Inf Q 2020 Jul;37(3):101481 [FREE Full text] [doi: 10.1016/j.giq.2020.101481] [Medline: 32313361]

94. Abbas A, Abdelsamea M, Gaber M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. medRxiv Preprint posted online on May 18, 2020. [doi: 10.1101/2020.03.30.20047456]

95. Ardabili S, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. COVID-19 Outbreak Prediction with Machine Learning. medRxiv Preprint posted online on April 22, 2020. [doi: 10.1101/2020.04.17.20070094]

96. Zhang H, Dimitrov D, Simpson L, Singh B, Plaks N, Penney S, et al. A Web-based, Mobile Responsive Application to Screen Healthcare Workers for COVID Symptoms: Descriptive Study. medRxiv Preprint posted online on April 22, 2020. [doi: 10.1101/2020.04.17.20069211]

97. Hart JL, Turnbull AE, Oppenheim IM, Courtright KR. Family-Centered Care During the COVID-19 Era. J Pain Symptom Manage 2020 Aug;60(2):e93-e97 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.04.017] [Medline: 32333961]

98. Parikh A, Kumar AA, Jahangir E. Cardio-Oncology Care in the Time of COVID-19 and the Role of Telehealth. JACC CardioOncol 2020 Jun;2(2):356-358 [FREE Full text] [doi: 10.1016/j.jaccao.2020.04.003] [Medline: 32328590]

XSL•FO
RenderX

99. Rahman MS, Peeri NC, Shrestha N, Zaki R, Haque U, Hamid SHA. Defending against the Novel Coronavirus (COVID-19) outbreak: How can the Internet of Things (IoT) help to save the world? Health Policy Technol 2020 Jun;9(2):136-138 [FREE Full text] [doi: 10.1016/j.hlpt.2020.04.005] [Medline: 32322475]

100. Alwashmi MF. The Use of Digital Health in the Detection and Management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2906 [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]

101. Sedov L, Krasnochub A, Polishchuk V. Modeling quarantine during epidemics and mass-testing using drones. medRxiv Preprint posted online on April 23, 2020. [doi: 10.1101/2020.04.15.20067025]

102. Mahapatra S, Nath P, Chatterjee M. Repurposing Therapeutics for COVID-19: Rapid Prediction of Commercially available drugs through Machine Learning and Docking. medRxiv Preprint posted online on April 23, 2020. [doi: 10.1101/2020.04.05.20054254]

103. Azizy A, Fayaz M, Agirbasli M. Do Not Forget Afghanistan in Times of COVID-19: Telemedicine and the Internet of Things to Strengthen Planetary Health Systems. OMICS 2020 Jun;24(6):311-313. [doi: 10.1089/omi.2020.0053] [Medline: 32324451]

104. Madurai Elavarasan R, Pugazhendhi R. Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic. Sci Total Environ 2020 Jul 10;725:138858 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138858] [Medline: 32336562]

105. Park S, Choi GJ, Ko H. Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea-Privacy Controversies. JAMA 2020 Jun 02;323(21):2129-2130. [doi: 10.1001/jama.2020.6602] [Medline: 32324202]

106. Negrini S, Kiekens C, Bernetti A, Capecci M, Ceravolo MG, Lavezzi S, et al. Telemedicine from research to practice during the pandemic. Eur J Phys Rehabil Med 2020 Apr 24:327-330 [FREE Full text] [doi: 10.23736/S1973-9087.20.06331-5] [Medline: 32329593]

107. Tanaka MJ, Oh LS, Martin SD, Berkson EM. Telemedicine in the Era of COVID-19: The Virtual Orthopaedic Examination. J Bone Joint Surg Am 2020 Jun 17;102(12):e57 [FREE Full text] [doi: 10.2106/JBJS.20.00609] [Medline: 32341311]

108. Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS One 2020;15(4):e0232391 [FREE Full text] [doi: 10.1371/journal.pone.0232391] [Medline: 32330208]

109. Javaid M, Haleem A, Vaishya R, Bahl S, Suman R, Vaish A. Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):419-422 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.032] [Medline: 32344370]

110. Kyhlstedt M, Andersson SW. Diagnostic and digital solutions to address the COVID-19 pandemic: The need for international collaboration to close the gap. Health Policy Technol 2020 Jun;9(2):126-128 [FREE Full text] [doi: 10.1016/j.hlpt.2020.04.010] [Medline: 32341908]

111. Barbosa RDM, Fernandes MAC. Chaos game representation dataset of SARS-CoV-2 genome. Data Brief 2020 Jun;30:105618 [FREE Full text] [doi: 10.1016/j.dib.2020.105618] [Medline: 32341946]

112. Blake H, Bermingham F, Johnson G, Tabner A. Mitigating the Psychological Impact of COVID-19 on Healthcare Workers: A Digital Learning Package. Int J Environ Res Public Health 2020 Apr 26;17(9):2997 [FREE Full text] [doi: 10.3390/ijerph17092997] [Medline: 32357424]

113. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]

114. Khan F, Ghaffar A, Khan N, Cho SH. An Overview of Signal Processing Techniques for Remote Health Monitoring Using Impulse Radio UWB Transceiver. Sensors (Basel) 2020 Apr 27;20(9):2479 [FREE Full text] [doi: 10.3390/s20092479] [Medline: 32349382]

115. Whelan P, Stockton-Powdrell C, Jardine J, Sainsbury J. Comment on "Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow": A UK Perspective. JMIR Ment Health 2020 Apr 27;7(4):e19547 [FREE Full text] [doi: 10.2196/19547] [Medline: 32330113]

116. Meinert E, Milne-Ives M, Surodina S, Lam C. Agile Requirements Engineering and Software Planning for a Digital Health Platform to Engage the Effects of Isolation Caused by Social Distancing: Case Study. JMIR Public Health Surveill 2020 May 06;6(2):e19297 [FREE Full text] [doi: 10.2196/19297] [Medline: 32348293]

117. Ekong I, Chukwu E, Chukwu M. COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria. JMIR mHealth uHealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]

118. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]

119. Kim H, Paul A. Contact Tracing: a game of big numbers in the time of COVID-19. medRxiv Preprint posted online on September 13, 2020. [doi: 10.1101/2020.04.22.20071043]

XSL•FO
RenderX

120. Krukowski RA, Ross KM. Measuring Weight with Electronic Scales in Clinical and Research Settings During the Coronavirus Disease 2019 Pandemic. Obesity (Silver Spring) 2020 Jul;28(7):1182-1183 [FREE Full text] [doi: 10.1002/oby.22851] [Medline: 32339394]

121. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]

122. Vaid A, Somani S, Russak A, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in COVID-19 Positive New York City Patients. medRxiv Preprint posted online on April 28, 2020. [doi: 10.1101/2020.04.26.20073411]

123. Lee A. COVID-19 and the Advancement of Digital Physical Therapist Practice and Telehealth. Phys Ther 2020 Jul 19;100(7):1054-1057 [FREE Full text] [doi: 10.1093/ptj/pzaa079] [Medline: 32343836]

124. Tarek M, Elhefnawi M, Maricato J, Diaz RS, Shytaj IL, Savarino A. Custommune: a web tool to design personalized and population-targeted vaccine epitopes. medRxiv Preprint posted online on April 29, 2020. [doi: 10.1101/2020.04.25.20079426]

125. Awasthi R, Pal R, Singh P, Nagori A, Reddy S, Gulati A, et al. CovidNLP: A Web Application for Distilling Systemic Implications of COVID-19 Pandemic with Natural Language Processing. medRxiv Preprint posted online on April 29, 2020. [doi: 10.1101/2020.04.25.20079129]

126. Khan H, Hossain A. Countries are Clustered but Number of Tests is not Vital to Predict Global COVID-19 Confirmed Cases: A Machine Learning Approach. medRxiv Preprint posted online on April 20, 2020. [doi: 10.1101/2020.04.24.20078238]

127. Husnayain A, Shim E, Fuad A, Su E. Assessing the community risk perception toward COVID-19 outbreak in South Koreavidence from Google and NAVER relative search volume. medRxiv Preprint posted online on April 29, 2020. [doi: 10.1101/2020.04.23.20077552]

128. Weemaes M, Martens S, Cuypers L, Van Elslande J, Hoet K, Welkenhuysen J, et al. Laboratory information system requirements to manage the COVID-19 pandemic: A report from the Belgian national reference testing center. J Am Med Inform Assoc 2020 Aug 01;27(8):1293-1299 [FREE Full text] [doi: 10.1093/jamia/ocaa081] [Medline: 32348469]

129. Espinoza J, Crown K, Kulkarni O. A Guide to Chatbots for COVID-19 Screening at Pediatric Health Care Facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]

130. Wagner T, Shweta F, Murugadoss K, Awasthi S, Venkatakrishnan AJ, Bade S, et al. Augmented Curation of Unstructured Clinical Notes from a Massive EHR System Reveals Specific Phenotypic Signature of Impending COVID-19 Diagnosis. medRxiv Preprint posted online on June 11, 2020. [doi: 10.1101/2020.04.19.20067660]

131. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, APHP/Universities/INSERM COVID-19 research collaboration, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), et al. International Electronic Health Record-Derived COVID-19 Clinical Course Profiles: The 4CE Consortium. medRxiv Preprint posted online on June 08, 2020. [FREE Full text] [doi: 10.1101/2020.04.13.20059691]

132. Hegde C, Jiang Z, Suresha P, Zelko J, Seyedi S, Smith MA, et al. AutoTriage - An Open Source Edge Computing Raspberry Pi-based Clinical Screening System. medRxiv Preprint posted online on April 30, 2020. [doi: 10.1101/2020.04.09.20059840]

133. Tobias A, Valls J, Satorra P, Tebe C. COVID19-Tracker: A shiny app to produce comprehensive data visualization for SARS-CoV-2 epidemic in Spain. medRxiv Preprint posted online on April 30, 2020. [doi: 10.1101/2020.04.01.20049684]

134. Compton M, Soper M, Reilly B, Gettle L, List R, Bailey M, et al. A Feasibility Study of Urgent Implementation of Cystic Fibrosis Multidisciplinary Telemedicine Clinic in the Face of COVID-19 Pandemic: Single-Center Experience. Telemed J E Health 2020 Aug;26(8):978-984. [doi: 10.1089/tmj.2020.0091] [Medline: 32357084]

135. Smith WR, Atala AJ, Terlecki RP, Kelly EE, Matthews CA. Implementation Guide for Rapid Integration of an Outpatient Telemedicine Program During the COVID-19 Pandemic. J Am Coll Surg 2020 Aug;231(2):216-222.e2 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2020.04.030] [Medline: 32360960]

136. Kalteh EA, Rajabi A. COVID-19 and digital epidemiology. Z Gesundh Wiss 2020 Apr 30:1-3 [FREE Full text] [doi: 10.1007/s10389-020-01295-y] [Medline: 32355606]

137. Woo Baidal JA, Chang J, Hulse E, Turetsky R, Parkinson K, Rausch JC. Zooming Toward a Telehealth Solution for Vulnerable Children with Obesity During Coronavirus Disease 2019. Obesity (Silver Spring) 2020 Jul;28(7):1184-1186 [FREE Full text] [doi: 10.1002/oby.22860] [Medline: 32352650]

138. Ciaffi J, Meliconi R, Landini M, Ursini F. Google trends and COVID-19 in Italy: could we brace for impact? Intern Emerg Med 2020 May 25:1555-1559 [FREE Full text] [doi: 10.1007/s11739-020-02371-7] [Medline: 32451932]

139. Carullo G. Digitalization of controls at the time of coronavirus. Article in Italian. CERIDAP. 2020 Mar 31. URL: https://ceridap.eu/digitalizzazione-dei-controlli-ai-tempi-del-coronavirus/ [accessed 2020-11-04]

140. Article 25. Data protection by design and by default. GDPR Text. URL: https://gdpr-text.com/read/article-25/ [accessed 2020-11-04]

141. Doran CM, Kinchin I. Economics of Mental Health: Providing a Platform for Efficient Mental Health Policy. Appl Health Econ Health Policy 2020 Apr;18(2):143-145. [doi: 10.1007/s40258-020-00569-6] [Medline: 32173795]

142. Sweeney Y. Tracking the debate on COVID-19 surveillance tools. Nat Mach Intell 2020 Jun 16;2(6):301-304. [doi: 10.1038/s42256-020-0194-1]

143. Wang S, Ding S, Xiong L. A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. JMIR mHealth uHealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]

144. Garg S, Bhatnagar N, Gangadharan N. A Case for Participatory Disease Surveillance of the COVID-19 Pandemic in India. JMIR Public Health Surveill 2020 Apr 16;6(2):e18795 [FREE Full text] [doi: 10.2196/18795] [Medline: 32287038]

145. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]

146. Corrao G, Rea F, Di Martino M, Lallo A, Davoli M, DlE Pl A, working group 'Monitoring and assessing diagnostic-therapeutic paths' of the Italian Heath Ministry. Effectiveness of adherence to recommended clinical examinations of diabetic patients in preventing diabetes-related hospitalizations. Int J Qual Health Care 2019 Jul 01;31(6):464-472. [doi: 10.1093/intqhc/mzy186] [Medline: 30204865]

147. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]

148. De Groote SL, Shultz M, Blecic DD. Information-seeking behavior and the use of online resources: a snapshot of current health sciences faculty. J Med Libr Assoc 2014 Jul 17;102(3):169-176 [FREE Full text] [doi: 10.3163/1536-5050.102.3.006] [Medline: 25031557]

149. Callaway E. Will the pandemic permanently alter scientific publishing? Nature 2020 Jun;582(7811):167-168. [doi: 10.1038/d41586-020-01520-4] [Medline: 32504015]

Abbreviations

AI: artificial intelligence
CT: computed tomography
EHR: electronic health record
ePPE: electronic personal protective equipment
EXPH: Expert Panel on Effective Ways of Investigating in Health
GDPR: general data protection regulation
IoT: Internet of Things
IP: influenza pneumonia
IT: information technology
NCP: novel coronavirus pneumonia
PRISMA: Preferred Reporting Items for Systematic Reviews
QR code: quick response code

Original Paper

COVID-19 Outcome Prediction and Monitoring Solution for Military Hospitals in South Korea: Development and Evaluation of an Application

JoonNyung Heo[1*], MD; Ji Ae Park[2*], MS; Deokjae Han[3], MD; Hyung-Jun Kim[3], MD; Daeun Ahn[4], BS; Beomman Ha[1], MD, PhD; Woong Seog[1], MD; Yu Rang Park[2], PhD

[1]Armed Forces Medical Command, Seongnam, Republic of Korea

[2]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Internal Medicine, The Armed Forces Capitol Hospital, Seongnam, Republic of Korea

[4]Department of Nursing, The Armed Forces Capitol Hospital, Seongnam, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2493
Email: yurangpark@yuhs.ac

Abstract

Background: COVID-19 has officially been declared as a pandemic, and the spread of the virus is placing sustained demands on public health systems. There are speculations that the COVID-19 mortality differences between regions are due to the disparities in the availability of medical resources. Therefore, the selection of patients for diagnosis and treatment is essential in this situation. Military personnel are especially at risk for infectious diseases; thus, patient selection with an evidence-based prognostic model is critical for them.

Objective: This study aims to assess the usability of a novel platform used in the military hospitals in Korea to gather data and deploy patient selection solutions for COVID-19.

Methods: The platform's structure was developed to provide users with prediction results and to use the data to enhance the prediction models. Two applications were developed: a patient's application and a physician's application. The primary outcome was requiring an oxygen supplement. The outcome prediction model was developed with patients from four centers. A Cox proportional hazards model was developed. The outcome of the model for the patient's application was the length of time from the date of hospitalization to the date of the first oxygen supplement use. The demographic characteristics, past history, patient symptoms, social history, and body temperature were considered as risk factors. A usability study with the Post-Study System Usability Questionnaire (PSSUQ) was conducted on the physician's application on 50 physicians.

Results: The patient's application and physician's application were deployed on the web for wider availability. A total of 246 patients from four centers were used to develop the outcome prediction model. A small percentage (n=18, 7.32%) of the patients needed professional care. The variables included in the developed prediction model were age; body temperature; predisease physical status; history of cardiovascular disease; hypertension; visit to a region with an outbreak; and symptoms of chills, feverishness, dyspnea, and lethargy. The overall C statistic was 0.963 (95% CI 0.936-0.99), and the time-dependent area under the receiver operating characteristic curve ranged from 0.976 at day 3 to 0.979 at day 9. The usability of the physician's application was good, with an overall average of the responses to the PSSUQ being 2.2 (SD 1.1).

Conclusions: The platform introduced in this study enables evidence-based patient selection in an effortless and timely manner, which is critical in the military. With a well-designed user experience and an accurate prediction model, this platform may help save lives and contain the spread of the novel virus, COVID-19.

XSL•FO
RenderX

Introduction

COVID-19, which is caused by the SARS-CoV-2 virus, has officially been declared a pandemic [1]. Despite the effort, there were still casualties caused by the disease. The mortality rate of the disease varies between countries and even between cities in China [2]. There are speculations that these mortality rate discrepancies between regions are due to the availability of medical resources in that area. For instance, the mortality rate difference between the city of Hubei and other provinces is significant (~3% for Hubei and 0.7% for the rest of China), and plotting the cases per population against mortality rates shows a positive correlation [2].

The spread of the virus is currently placing sustained demands on public health systems [3]. At present, there is an imbalance in the supply and demand of medical supplies, and many efforts are being made to solve this problem [4]. Although it is ideal if the increase in medical supplies could meet the high demand, this seems to be difficult for many underdeveloped countries.

Adjusting demand, such as by selecting the patients for diagnosis and treatment, can also be an option considering that the rates for severe cases are relatively low, with severe cases at 14% and critical cases at 5% based on Chinese reports [5]. Comparably, rates of the asymptomatic cases are surprisingly high, with some reporting up to 75% [6-8]. In this context, the Centers for Disease Control and Prevention (CDC) recommends that not all patients be hospitalized, considering the insufficient medical infrastructure. However, there is no definite guideline regarding patient selection due to insufficient evidence on which patients need professional care.

Military personnel are especially at risk, since the viral spread is the most critical in settings where people are living closely in groups [9,10]. The adenovirus infection is one of the examples of an infectious disease that is easily spread in the military. Additionally, soldiers are frequently relocated worldwide, which may accelerate the spread of the disease. For instance, in 2007, basic military trainees were infected with the adenovirus and were sent to multiple sites throughout the world, causing an outbreak in those military bases [9].

The confusion, urea, respiratory rate, blood pressure, and 65 years or older (CURB-65) score is a widely known scoring system for predicting the outcome of patients with pneumonia [11]. The presence of confusion, blood urea nitrogen, respiratory rate, systolic and diastolic blood pressure, and age are assessed to calculate a score ranging from 0 to 6, with 6 indicating the highest severity. The American Thoracic Society guidelines suggest that patients with CURB-65 scores of less than 2 are at a low risk of death and may be managed as outpatients. In response to the outbreak, this score was used for patients with COVID-19 [12]. However, even though patients with higher CURB-65 scores were more likely to be severe, more than 20% of the patients with lower scores also had poor outcomes, suggesting the need for a new system to guide patient selection.

Previous studies have developed prognostic models to predict mortality risk, hospital stay, and progression to severe state [13-21]. The largest sample size of these studies was 577 patients from China [19]. Ideally, prediction models should be based on a large data set covering multiple nations and races for vast adaptation. Four of these studies presented models (eg, decision tree, nomogram, and scoring rule) that could be used in clinical practice [14,16,17,21].

When a prediction model is being developed and deployed in the real world, multiple factors should be considered. As previously stated, the included samples should be from multiple centers across multiple nations to account for disparities between races. Additionally, the model should provide a practical and irreplaceable value to its users. The objective of the model should be clear and should be able to answer a major question in the clinical field. The model's ease of use should also be considered. However, these requirements are difficult—if not impossible—to meet in a situation where a novel infection is spreading worldwide.

In this study, we present a platform that provides outcome prediction and status monitoring for patients with COVID-19 that is consistently enhanced with data collected based on the use of the model.

Methods

Development of the Platform Structure

The platform structure was designed to primarily serve the users by providing the results of the prediction models and sending the results to the appointed physician. The data sent by the user to acquire the results were collected by the server with the user's consent. The newly collected data was used for training the models to further enhance its predictability and generalizability, thereby completing the virtuous cycle. There are two applications employed to serve this purpose: (1) a patient's application and (2) a physician's application.

The main functionality of the patient's application is to provide outcome prediction results to the patient with their general information, previously diagnosed diseases, symptoms, and body temperature. Specifically, the purpose of the outcome prediction is to guide patient selection for medical resource allocation. Thus, the main target of the prediction model is to determine whether the patient will need professional care.

The application receives the variables that the patient provides to acquire the prediction results, sends them to the central server, and shares them with the appointed physician for remote monitoring. When the patient visits a health care facility, the patient can pair their application with the physician's application to keep the physician notified of the change in the patient's

XSL•FO

RenderX

symptoms without needing any direct contact. The physician's application then receives a registration code that can be created from the patient's application. Upon registering a patient, the health care worker can receive all the records and prediction results from the patient's application. This process is depicted in Figure 1.

Figure 1. Diagram of the whole structure of the platform.
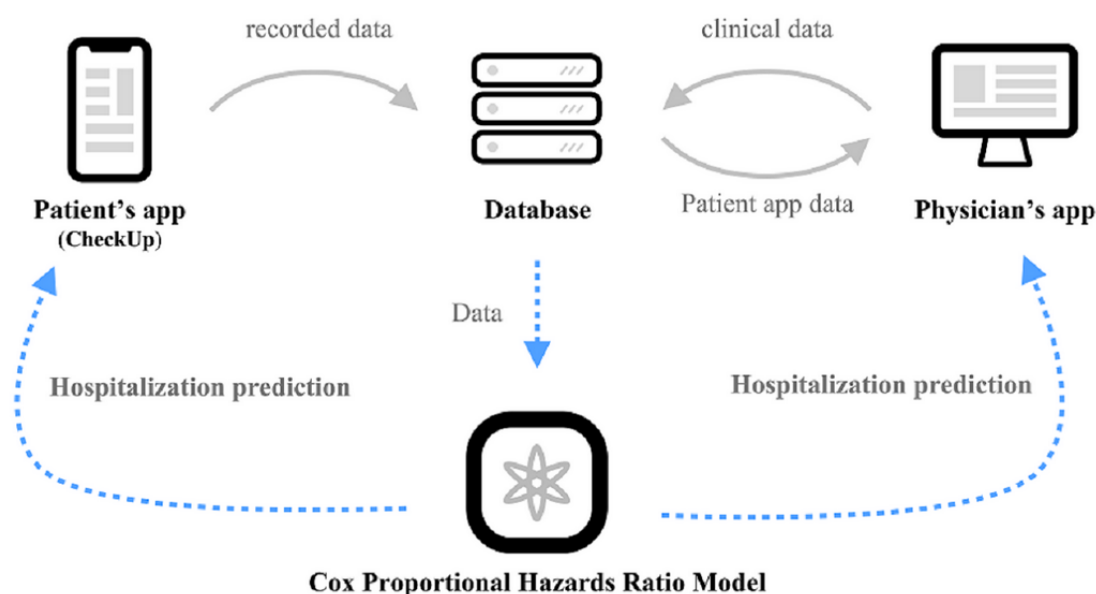
Three physicians reviewed previous publications related to COVID-19 to select the appropriate variables for collection [22-24]. On the basis that this novel virus might have similar features to the pre-existing viruses that cause viral pneumonia, the characteristics of other viral pneumonias were also considered. The selection result was then reviewed by two physicians who are currently directly involved with treating patients with COVID-19 (Multimedia Appendix 1).

Participants and Data Collection for Outcome Prediction

All patients with COVID-19 admitted to four centers (two military hospitals and two civilian hospitals) in Korea were included in the study. A total of 246 patients were included, and data collection ran from February 6, 2020, to April 2, 2020. There were no exclusion criteria.

The candidate risk factors for the patient's application included patient demographic characteristics (age, gender, and BMI), past history (asthma, chemotherapy, chronic kidney disease, hypertension, chronic obstructive pulmonary disease, immunosuppressant use, cardiovascular disease, chronic liver disease, and diabetes), patient symptoms (anosmia, rhinorrhea, chest pain, phlegm, chills, physical status, cough, pneumonia, diarrhea, pneumonia, antipyretic, dyspnea, feverish, headache, muscle pain, nausea or vomiting, tired or lethargy, and sore throat), social history (visit to a region with an outbreak, direct contact with a patient with COVID-19, smoking history, household member confirmed with COVID-19, and household member under self-isolation), and body temperature during hospitalization.

The outcome measured in this study was the use of an oxygen supplement monitored up to April 6, 2020—the final date of follow-up. The length of time (days) from the date of hospitalization to the date of first use of an oxygen supplement was considered as our target outcome. Oxygen supplement was chosen as the outcome measure since it may be able to represent the minimal treatment required for hospital admission.

Statistical Analysis

Descriptive statistics were obtained for all study variables. Continuous data were expressed as mean (SD) values, and categorical data were expressed as proportions.

A survival curve for overall patients was plotted using the Kaplan-Meier method. Using all candidate risk factors, we conducted a univariate analysis by a Cox proportional hazards model to select factors. The factors with $P<.05$ were included in the final model. The body temperature collected during hospitalization was used as the time-dependent variable in the Cox proportion hazards model [25]. By generating time intervals during the day after the date of hospitalization, the average body temperature recorded at each time interval (day) was calculated. That is, the body temperature was considered as a value that changed every day. In case of missing body temperature, the last observation carried forward approach was used. A multivariate Cox proportional hazard model (with time-dependent variables) using selected factors was considered as the prediction model for the patient's application. The multicollinearity between the factors included in the final model was confirmed through variance inflation factors (VIFs), which quantify the severity of multicollinearity in an ordinary least squares regression analysis. We considered that VIF of 5 or above indicates a multicollinearity problem [26]. Also, additionally, association between factors was examined through phi coefficient between two binary factors, point-biserial correlation between a continuous and a binary factor, and pearson correlation between two continuous factors. To calculate VIF through the ordinary least square regression analysis and

to measure correlation between factors, the median body temperature during hospitalization was considered for body temperature with time-varying characteristics.

We calculated the Harrell C statistic to assess the overall predictive accuracy and time-dependent area under the receiver operating characteristic curve (area under the curve [AUC]) [27] to assess the predictive accuracy over the entire follow-up period. The AUC results of the Cox model, which vary depending on the time point to be evaluated, were summarized by days. We also performed the likelihood ratio, Wald, and score tests to measure overall goodness of fit based on the omnibus test of model coefficients. The chi-square goodness-of-fit test proposed by Schoenfeld [28] was used for the proportional hazards. Since body temperature was considered as a time-varying variable updated every day in the model, we did not check the assumption for body temperature. A random subsampling [29] was used to estimate the internal validity of the final Cox proportional hazards model, with 50 repeated samplings dividing the train and test sets into 2:1. The sampling was performed by stratifying according to the occurrence of the outcome event to establish balance in the test and train sets. An average time-dependent AUC based on the 50 repetitions was used as a result of the validation.

The Cox proportional hazards model was considered as a prediction model for the patient's application, but we additionally constructed a logistic model as an alternative model by using selected factors from the final Cox model. The logistic model was designed to predict oxygen supplement use during hospitalization.

All P values were two-sided, and $P<.05$ was considered statistically significant. Statistical analysis was performed using R 3.6.0 (R Foundation for Statistical Computing) [29].

Evaluation of the Physician's Application

A usability study was conducted to evaluate the usability of the health care worker's application wherein participants had to be medical doctors with more than 1 year of clinical practice. A total of 50 physicians were recruited from a public website, and there was no exclusion or selection of the participants. The participants were introduced to the entire platform and briefed

about its objectives. After a thorough explanation of the study, a consent form was signed by the participants. They were then instructed to sign up for the service through the physician's application and then add a prepared sample patient case that was provided. Afterward, participants were required to complete several tasks: review the symptoms that the sample case reported through the patient's application, review the results from the outcome prediction model of the registered patient, and add additional factitious clinical variables for the patient from the variable input form. The result from the model was presented as the probability of requiring an oxygen supplement, expressed as a number between 0%-100%. The participants were educated on the intended clinical utility of the result from the prediction model to provide supportive information that can be used during patient selection. However, the usefulness of the prediction model could not be assessed since the data were not from a real patient. The participants were instructed to fill the factitious values of the patients' clinical variables required by the physician's input form. The study protocol was approved by the institutional review board at the Yonsei University College of Medicine (4-2020-0351).

Following the completion of the tasks, the Post-Study System Usability Questionnaire (PSSUQ) was used to acquire the participants' responses [30]. The results from the PSSUQ were analyzed using statistical representative indexes for each question.

Results

Platform Structure

The platform consists of three parts: (1) the patient's application, (2) the physician's application, and (3) the model application programming interface. Both the patient's and the physician's applications are deployed on the web. The web platform was chosen to be able to enable instant modification upon deployment and maximize the availability on variable devices. Considering the majority of devices the users will be using, the patient's application is configured to be best viewed on a mobile device and the physician's application on a desktop. The diagram of the entire structure is illustrated in Figure 2.

Figure 2. Screenshots showing how patient recorded data is shown in the physician's app to monitor the patient's status.
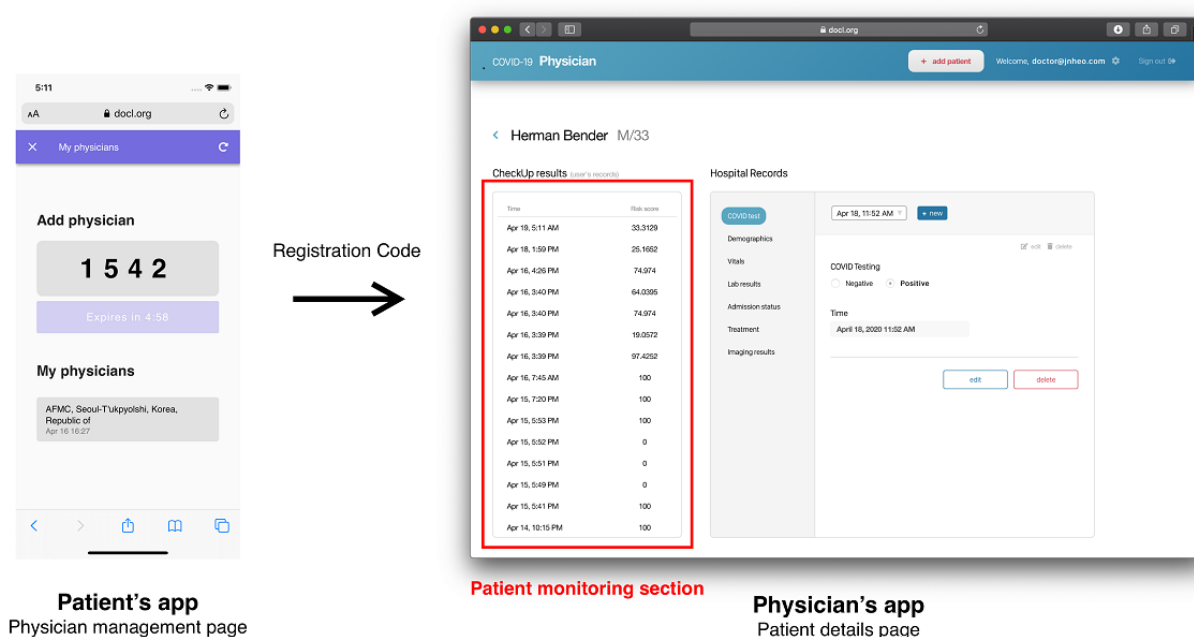
This platform is registered in the World Health Organization Digital Atlas for COVID-19 solutions [31].

Real-World Use of the Platform

A patient is advised to use the patient's application when they are officially diagnosed with COVID-19. They are also advised to use the application every day and provide information about their daily symptoms and body temperature. If the result from the application shows a high likelihood of requiring hospitalization and the patient is not currently admitted to a hospital, then the patient is advised to contact a nearby health care professional. If there is a previously appointed physician monitoring the patient's daily symptoms, then the physician can check the daily status of the patient. Using the results of the application, the physician is then able to make a proper clinical decision about whether the patient will need hospitalization. The result that the physician receives is the raw probability calculated from the prediction model. No cut-off was set, since the cut-off will be affected by environmental factors, such as the availability of hospital beds or the regional number of confirmed patients.

Development of the Outcome Prediction Model

The baseline characteristics of all 246 patients are shown in Multimedia Appendix 2. The mean (SD) of age (year) and BMI (kg/m^2) were 40.72 (SD 17.10) years and 3.21 (SD 3.28) kg/m^2, respectively. A total of 167 (67.89%) patients were male. The mean (SD) of the max and median body temperature during

follow-up were 37.32 °C (SD 0.56) and 36.75 °C (SD 0.42), respectively. The most common symptom was coughing (95 patients, 41.13%), and 192 (83.12%) patients visited a region with a COVID-19 outbreak.

One patient with missing date of hospitalization among 246 patients was excluded to estimate the survival rate for oxygen supplement. During 2469.32 person-days of follow-up, oxygen supplement was observed in 18 (7.35%) patients. The median (IQR) time from hospitalization to censoring was 17.64 (IQR 14.24-21.70) days. The Kaplan-Meier estimates of the 3-day and 5-day survival rate were 0.947 (95% CI 0.919-0.975) and 0.930 (95% CI 0.898-0.963), respectively.

Table 1 shows results of the univariate and multivariate Cox proportional hazards models. In the univariate analysis, 10 of the candidate risk factors showed results of $P<.05$ (Table 1). It was confirmed that hypertension (hazard ratio [HR] 3.792, 95% CI 1.423-10.109), cardiovascular disease (HR 12.413, 95% CI 4.069-37.869), predisease physical status (HR 1.854, 95% CI 1.24-2.773), dyspnea (HR 13.498, 95% CI 4.527-40.252), feeling feverish (HR 6.282, 95% CI 2.054-19.213), chills (HR 5.727, 95% CI 1.924-17.048), tiredness or lethargy (HR 6.083, 95% CI 1.989-18.607), older age (HR 1.075, 95% CI 1.052-1.098), and higher body temperature (HR 13.147, 95% CI 6.849-25.237) were risk factors for earlier oxygen supplement, while visiting a region with an outbreak (HR 0.291, 95% CI 0.095-0.89) was confirmed as a risk factor that decreased the risk of oxygen supplement.

Table 1. Hazard ratio of univariate and multivariate Cox proportional hazards model.

Factors	Univariate model, HR[a] (95% CI)	Multivariate model[b]	
		HR (95% CI)	P value
Age (years)	1.075 (1.052-1.098)	1.035 (0.983-1.089)	.47
Body temperature[b] (°C)	13.147 (6.849-25.237)	17.431 (2.856-106.371)	.01
Hypertension (yes)	3.793 (1.423-10.109)	0.562 (0.033-9.437)	.81
CVD[c] (yes)	12.413 (4.069-37.869)	0.217 (0.008-6.111)	.49
Visit to a region of outbreak (yes)	0.291 (0.095-0.89)	3.381 (0.133-86.084)	.28
Physical status	1.854 (1.24-2.773)	4.259 (1.679-10.802)	.007
Dyspnea (yes)	13.498 (4.527-40.252)	3.878 (0.454-33.111)	.43
Feverish (yes)	6.282 (2.054-19.213)	0.321 (0.05-2.073)	.26
Chills (yes)	5.727 (1.924-17.048)	0.905 (0.049-16.705)	.95
Tired/lethargic (yes)	6.083 (1.989-18.607)	1.506 (0.174-13.019)	.62

[a]HR: hazard ratio.

[b]Cox proportional hazards model with time-dependent variable.

[c]CVD: cardiovascular disease.

As a result of multivariate analysis including all 10 factors, body temperature (HR 17.431, 95% CI 2.856-106.371) was found to be the most powerful risk factor for oxygen supplement. Physical status remained statistically significant as a factor (P=.007), while the other eight factors were not statistically significant.

The proportional hazards assumption on the factors included in the multivariate model (except for body temperature, since it is considered as a time-varying variable) was checked. Results showed that all factors satisfied the assumption (Multimedia Appendix 3). The multicollinearity and association between the factors included in the multivariate model were not detected (Multimedia Appendices 4 and 5).

To confirm the effects of body temperature in the multivariate model, we considered a case for a fully active man 40 years of age without any past history or symptoms and compared survival rates according to body temperature (Figure 3). In cases of temperatures of 37 °C, 38 °C, 38.5 °C, and 39 °C, the 5-day survival rates were 0.994 (95% CI 0.9810-1), 0.906 (95% CI 0.7071-1), 0.663 (95% CI 0.1782-1), and 0.180 (95% CI 0.0002-1), respectively. At 39 °C, the 5-day survival rate rapidly decreased.

Figure 3. Survival rate according to body temperature in predictive model present in the patient's application. The 5-day survival rates for each initial body temperature are shown.
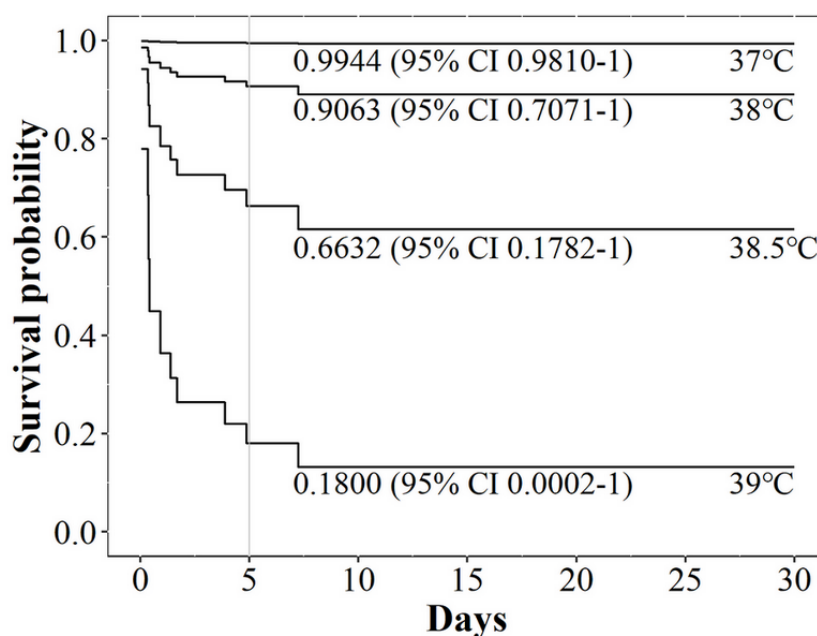
For the multivariate model, the overall C statistic was 0.963 (95% CI 0.936-0.99), and the time-dependent AUC ranged from 0.976 at day 3 to 0.979 at day 9 (Figure 4). Multimedia Appendix 6 shows the time-dependent AUC ranging from day

1 to day 10. The value for each day was above 0.96. The AUC of 8 days or more were the same because the outcome event did not occur after 7.25 days.

Figure 4. Time-dependent area under the receiver operating characteristic curve at 3, 6, and 9 days from the prediction model present in the patient's application.
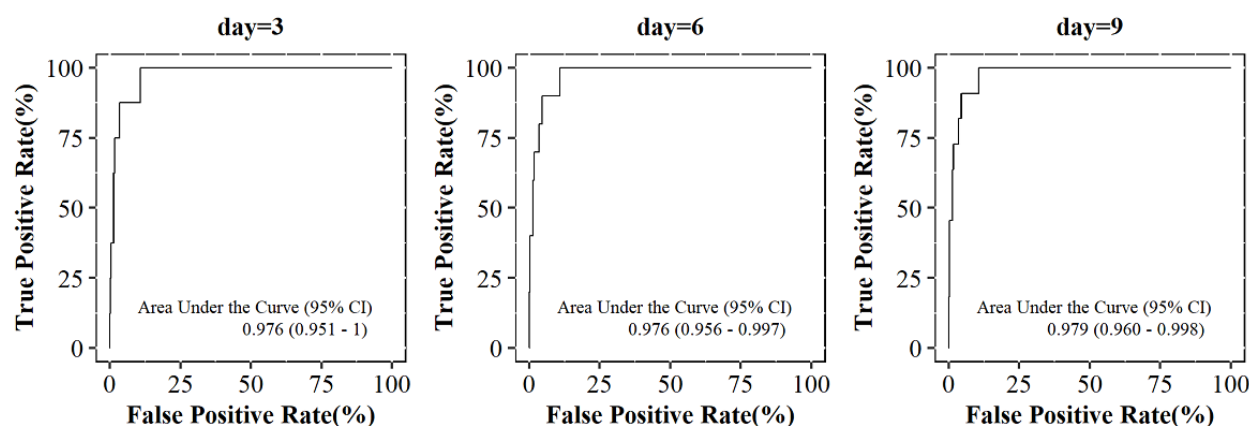
As per the result of internal validation for estimating the performance and validity of the multivariate Cox proportional hazards model in Table 1, the average time-dependent AUC ranging from day 1 to day 10 showed a range from a minimum of 0.749 to a maximum of 0.764 (Multimedia Appendix 7).

The results from the model are displayed at the result section of the application when the user has completed a series of questionnaires. The results page shows the 7day survival result based on the received variables.

As results of the multivariate logistic model, body temperature (odds ratio [OR] 19.106, 95% CI 1.587-229.961), physical status (OR 5.145, 95% CI 1.539-17.205), and age (OR 1.099, 95% CI 1.028-1.174) were statistically associated with oxygen supplementation during hospitalization (Multimedia Appendix 8). The P values of the likelihood ratio and score tests were less than .001, and the Hosmer and Lemeshow test showed a P value

of .99. In addition, the AUC was 0.973 (95% CI 0.9413-1; Multimedia Appendix 9).

Platform Evaluation

The usability study was performed with 50 participants from May 1, 2020, to May 31, 2020. Participants had a mean age of 35.9 years, and 35 (70.0%) were men. All participants were able to complete the given tasks easily. The detailed results are shown in Table 2. The participants were instructed to answer the PSSUQ on a scale of 1-7, with "1" being the most agreeable to the statement. The lower the values of the answers, the better the user experience was. Overall, the participants were satisfied with the app, with the overall average of the responses being 2.2 (SD 1.1). The most disagreeable statement from the questionnaire was the seventh statement ("The system gave error messages that clearly told me how to fix problems"), which had a mean response of 2.7. However, no error was experienced during the study process, thus many of the participants chose "4" as a response, which stood for "neutral."

Table 2. Usability study results (score ranging from 1 to 7, one being "strongly agree").

Statements	Score, mean (SD)
1. Overall, I am satisfied with how easy it is to use this system.	2.0 (1.0)
2. It was simple to use this system.	2.0 (1.1)
3. I was able to complete the tasks and scenarios quickly using this system.	2.2 (1.1)
4. I felt comfortable using this system.	2.2 (1.2)
5. It was easy to learn to use this system.	2.0 (1.1)
6. I believe I could become productive quickly using this system.	2.0 (1.1)
7. The system gave error messages that clearly told me how to fix problems.	2.7 (1.3)
8. Whenever I made a mistake using the system, I could recover easily and quickly.	2.3 (1.2)
9. The information such as online help, on-screen messages, and other documentation provided with this system was clear.	2.0 (1.0)
10. It was easy to find the information I needed.	2.3 (1.2)
11. The information was effective in helping me complete the tasks and scenarios.	2.1 (1.1)
12. The organization of information on the system screens was clear.	2.0 (1.1)
13. The interface of this system was pleasant.	2.1 (1.2)
14. I liked using the interface of this system.	2.1 (1.2)
15. This system has all the functions and capabilities I expect it to have.	2.4 (1.2)
16. Overall, I am satisfied with this system.	2.1 (1.0)

Discussion

Principal Findings

In this study, we developed a platform to provide the users with the most up to date, evidence-based prediction model that can guide them in making decisions on whether to seek professional care or hospitalization, using only the variables provided by the users. The result provided to the users is the probability of the patient requiring oxygen supplement. With this, patients can effectively and accurately monitor themselves on whether they will need hospitalization during self-quarantine. Physicians can use the result from the prediction model for patient selection for hospitalization and risk assessment when needed. The platform collects the data required to calculate the prediction results and uses the data to update the prediction results given directly to the users. The developed Cox model in this study has a high accuracy with an AUC of 0.97 or higher. Our model showed that body temperature was the most important factor for oxygen supplement. A usability study was performed with PSSUQ, which showed that the participants were generally satisfied with the application.

Timely Patient Selection and the Military

The platform is currently used in military hospitals in Korea. Collection of data and deployment of the outcome prediction model is essential in military settings. If a soldier were to be confirmed with the disease, an exceptionally large number of people would be at risk. In contrast, the patients are likely to be considered low risk, that is, having low probability of requiring professional care, since most of them will be younger and healthier than the general population. This special circumstance, where an explosive spread of the virus is expected but only a few will need special care, will pose a greater importance in the selection of patients for both diagnosis and treatment. A large fraction of the confirmed patients will not be hospitalized, so a close and efficient method for monitoring these patients is necessary. However, the platform was developed considering its use outside of the military when needed. Therefore, patients from two public hospitals were included in the study to address this possibility.

The platform is tailored to fit needs in situations like the COVID-19 pandemic, where data of the novel disease is scarce, and the disease is spreading so fast that traditional clinical trials are not timely enough. Clinical trials are carefully designed to minimize bias and clearly prove a hypothesis. Accuracy and reliability are critical in the medical field, but progressive measures may be required during desperate times. Furthermore, real-world data may be more representative of the patient populations in the clinical field [32]. However, the results derived from this platform should be used in a complementary manner, especially if there is more reliable evidence such as a randomized controlled trial.

The main goal of this platform is to provide the physician with a supportive measure to assess the patient's risk of requiring professional care. The probability of the patient requiring admission is provided to the physician in the form of a percentage (ie, 0%-100%). The model does not decide if the patient will need hospitalization but provides the risk information of the patient to the physician so a proper decision can be made. Considering the availability of medical resources and the number of patients confirmed with COVID-19, the physician will have to decide which patient will require hospitalization. The prediction model is expected to help the physician with complementary information for the better prioritization of patients.

Evaluation of the Outcome Prediction Model

The model presented in our study used 246 patients' data from four centers in Korea to predict whether a patient would need hospitalization during the course of the disease. The Cox proportional hazards ratio model was used to account for the time-dependent variable (ie, body temperature), and the model showed a high predictive accuracy (C statistic: 0.963; AUC at 9 days: 0.979). Ten features were selected in the model: age, body temperature, history of hypertension, cardiovascular disease, visit to a region with an outbreak, predisease physical status, dyspnea, feeling feverish, chills, and tiredness or lethargy. Body temperature showed the highest HR (17.431, 95% CI 2.856-106.371), followed by predisease physical status (HR 4.259, 95% CI 1.679-10.802) and dyspnea (HR 3.898, 95% CI 0.454-33.111; Table 1).

The predictors selected for the model are mostly consistent with previous reports. Although age and comorbidities are already well-established risk factors for grave outcomes [33,34], dyspnea was also found to be associated with disease progression [33]. In our study, the effect of body temperature was the most important predictor. For example, a patient who is a fully active man 40 years of age without any past history or symptoms was highly likely to need an oxygen supplement within 5 days with a 5-day survival rate of 0.18 at 39 °C (Figure 3). Abnormal body temperature is a well-known risk factor for grave prognosis in patients with community-acquired pneumonia [35]. Physical status, which is often expressed in the activities of daily living score, is known to be an independent risk factor for mortality among older adult patients with pneumonia [36,37].

When multivariable prediction models are developed, the number of outcome events compared to the number of predictors, referred to as the events per variable (EPV), affects the accuracy of the model. In general, it is known that a reliable sample size is at least 10 events per predictor (variable) in logistic and Cox models [38,39]. In our model, this rule is not satisfied for the 10 predictors included in the final model. The problem of this low EPV was affected by the accuracy and precision of the coefficients. In addition to the sample size, there is also a possibility that the standard error of the coefficient was overestimated due to the multicollinearity [40], and as a result of the VIF, there was no multicollinearity between predictors (Multimedia Appendix 10). Since overfitting most notably occurs when the number of candidate predictors is large relative to the number of outcome events [41], our model could also be an overfitted model. For this reason, the AUC in the total development sample was about 0.97 (Figure 4), but the result in the internal validation was about 0.75 (Multimedia Appendix 7).

Caution should be employed in the interpretation of several risk factors' HR in the developed Cox model. We constructed the multivariate Cox model by selecting only statistically significant variables from the univariate model. In the univariate model, the HRs of all risk factors were calculated in the expected direction for outcome risk. However, in the multivariate model, hypertension, cardiovascular disease, feeling feverish, and visits to a region with an outbreak were calculated in a different direction from the result in the univariate model. As these factors have a relatively weak influence compared to body temperature and physical status, and the statistical significance disappeared, caution should be taken in the interpretation of the estimated HRs in contrast to the univariate model.

In the developed Cox proportional hazards model, it is accurate to define the onset of follow-up as the date when the patient was infected with COVID-19. However, it was difficult to accurately estimate the date of infection. Therefore, the onset of follow-up was defined as the date of hospitalization or the date on which the patient first recorded data in the application. The inability to accurately estimate the period between the date when the patient was infected with COVID-19 and the date considered as the onset of follow-up can be considered as a limitation of the model.

Usability of the Physician's Application

Usability evaluation was performed for the physician's application since the use of the application should be effortless to lessen their overwhelming workload due to the COVID-19 pandemic. Additionally, the application will be used in the setting where a physician must monitor a large number of patients; therefore, the usability will have a major impact. The application is simple, presenting a list of associated patients on the first page and showing details of the patient when clicking on the list. Consequently, the use questionnaire showed promising results, with a mean score of the total PSSUQ being 2.2 (SD 1.1). The time taken to complete the given tasks was not recorded, but many participants had completed the tasks under 10 minutes and were surprised by how simple the tasks were. The most disagreeable question was the seventh question, which inquired whether the error message was easy to understand for the user. Since no participant experienced an error, the result from the question was unreliable.

Other Applications and Outcome Prediction Models for COVID-19

Although not evidence-based, there are some applications that help patients by providing behavioral guidance when they suspect COVID-19 infection. Our initially developed applications, CheckUp Classic and Triage Classic, provided expert opinion-based guidelines in patient selection for testing and patient triage, respectively. These applications received more than 240,000 visits worldwide in a month. Additionally, Apple Inc published a COVID-19 Screening Tool in cooperation with the CDC [42]. A self-triage and self-scheduling tool were also developed based on a well-designed algorithm. This tool is designed to be tethered to the electronic medical records [43].

A few models have been developed to predict the outcome for patients with COVID-19, and a systemic review has been performed [44]. In the review, the authors conclude that all of these models are not useful in the clinical setting. This is primarily due to poor adherence to guidelines and small data size, both of which are unavoidable in a situation where there is not enough time. In a recent editorial, the author emphasizes the importance of sharing data to overcome the limitations of building such a prediction model [45]. However, motive for the researchers to share the hard-earned patient data is weak. Data collection is an additional burden for the already occupied health

XSL•FO
RenderX

care providers. There are a few efforts to collect nationwide or worldwide data for patients with COVID-19 [46,47].

COVID-19 Estimated Risk (COVER) was one of the models developed using the common data model [48]. The model comprises nine variables: age, sex, history of cancer, chronic obstructive pulmonary disease, diabetes, heart disease, hypertension, hyperlipidemia, and kidney disease. It was developed based on 6,869,127 patients with influenza or flu-like symptoms. Although the score itself was not originally developed for patients with COVID-19, it was validated with 43,061 patients with COVID-19. The performance, measured with the AUC, was 0.73-0.81 for COVER-H (predicting hospitalization), 0.73-0.91 for COVER-I (predicting intensive care), and 0.82-0.90 for COVER-F (predicting fatality).

Future Research

Regardless of the accuracy or the representativeness of the data set used for the development of the prediction model, the clinical utility of the platform should be prospectively evaluated. Whether additional information from the prediction model is helpful for the physician or if it is redundant should be further studied. Considering that outcome may vary according to environmental factors, it is questionable if a single prediction model could be generalized worldwide or if there should be separate models for each environment. Currently the platform is not considered for evaluation as an authorized medical device. Extensive prospective research following collection of large data sets to fully represent the target population should be performed.

Limitations

There are limitations in this study. First, the design of the study does not conform to any previous guidelines for clinical studies.

This is due to the difference of scope and objective between traditional clinical studies and this platform. Even so, this study is prone to many types of bias, so the results from this study should be handled with care. Auditing of data is not possible since there is no direct contact with the users, so the completeness or accuracy of the data cannot be ensured. The platform itself is unique, so the study design has never been validated. Considering that the guideline for patient hospitalization is dependent on multiple factors including hospital capacity, governmental guidelines, and patient risk, a single outcome measure cannot be used. Thus, the primary outcome measure—oxygen supplement—is not a definite marker for hospital admission. However, for our data set, all cases were reviewed, and there was no patient that needed hospitalization without receiving oxygen supplement.

Our prediction model was developed from a multicenter data set. However, the model was constructed based on a small data set of a single ethnicity, which can cause selection bias. In addition, since the model has not been built with sufficient data size, there may be a problem in terms of the model's accuracy. With more data acquisition, we will reinforce our model to expand the population applied to the model and increase the generalizable possibilities by performing external validation for other patients.

Conclusion

The platform introduced in this study provides evidence-based decision support to guide patient selection in an effortless and timely manner, which is critical in the military. With a well-designed user experience and an accurate prediction model, this platform may help save lives and contain the spread of the novel virus, COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables collected for the patient's application.
[DOCX File , 18 KB - jmir_v22i11e22131_app1.docx]

Multimedia Appendix 2
Overall characteristics of patients at admission.
[DOCX File , 21 KB - jmir_v22i11e22131_app2.docx]

Multimedia Appendix 3
Chi-square test by using the Schoenfeld residuals for the proportional hazards.
[DOCX File , 15 KB - jmir_v22i11e22131_app3.docx]

XSL·FO

RenderX

Multimedia Appendix 4
Testing Global Null Hypothesis: Beta=0 of multivariate Cox proportional hazards model.
[DOCX File , 14 KB - jmir_v22i11e22131_app4.docx]

Multimedia Appendix 5
Correlation matrix between 10 predictor variables included in multivariate Cox proportional hazards model.
[DOCX File , 16 KB - jmir_v22i11e22131_app5.docx]

Multimedia Appendix 6
Time-dependent area under the receiver operating characteristic curve ranged from 1-day to 10-day of multivariate Cox proportional hazards model.
[DOCX File , 15 KB - jmir_v22i11e22131_app6.docx]

Multimedia Appendix 7
Average time-dependent area under the receiver operating characteristic curve ranged from 1-day to 10-day using 50 repeated random subsampling to estimate the internal validity.
[DOCX File , 14 KB - jmir_v22i11e22131_app7.docx]

Multimedia Appendix 8
Odds ratio of multivariate logistic model.
[DOCX File , 16 KB - jmir_v22i11e22131_app8.docx]

Multimedia Appendix 9
Results of overall model evaluation, goodness of fit, and predictive accuracy from multivariate logistic model.
[DOCX File , 16 KB - jmir_v22i11e22131_app9.docx]

Multimedia Appendix 10
Results of multicollinearity test with variance inflation factor.
[DOCX File , 15 KB - jmir_v22i11e22131_app10.docx]

References

1. WHO Timeline - COVID-19. World Health Organization. 2020 Apr 27. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-05-03]
2. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020 Apr;8(4):e480 [FREE Full text] [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]
3. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
4. Ranney ML, Griffeth V, Jha AK. Critical supply shortages — the need for ventilators and personal protective equipment during the COVID-19 pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/nejmp2006141]
5. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
6. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ 2020 Mar 23;368:m1165. [doi: 10.1136/bmj.m1165] [Medline: 32205334]
7. Wang Y, Liu Y, Liu L, Wang X, Luo N, Li L. Clinical outcomes in 55 patients with severe acute respiratory syndrome coronavirus 2 who were asymptomatic at hospital admission in Shenzhen, China. J Infect Dis 2020 May 11;221(11):1770-1774 [FREE Full text] [doi: 10.1093/infdis/jiaa119] [Medline: 32179910]
8. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci China Life Sci 2020 May;63(5):706-711 [FREE Full text] [doi: 10.1007/s11427-020-1661-4] [Medline: 32146694]
9. Trei JS, Johns NM, Garner JL, Noel LB, Ortman BV, Ensz KL, et al. Spread of adenovirus to geographically dispersed military installations, May-October 2007. Emerg Infect Dis 2010 May;16(5):769-775. [doi: 10.3201/eid1605.091633] [Medline: 20409365]
10. Kak V. Infections in confined spaces: cruise ships, military barracks, and college dormitories. Infect Dis Clin North Am 2007 Sep;21(3):773-84, ix [FREE Full text] [doi: 10.1016/j.idc.2007.06.004] [Medline: 17826623]

XSL•FO
RenderX

11. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003 May;58(5):377-382. [doi: 10.1136/thorax.58.5.377] [Medline: 12728155]

12. Nguyen Y, Corre F, Honsel V, Curac S, Zarrouk V, Fantin B, et al. Applicability of the CURB-65 pneumonia severity score for outpatient treatment of COVID-19. J Infect 2020 Sep;81(3):e96-e98 [FREE Full text] [doi: 10.1016/j.jinf.2020.05.049] [Medline: 32474039]

13. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020 Mar 18;24(1):108 [FREE Full text] [doi: 10.1186/s13054-020-2833-7] [Medline: 32188484]

14. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]

15. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, et al. Predicting COVID-19 malignant progression with AI techniques. SSRN J Preprint posted online on March 31, 2020. [doi: 10.2139/ssrn.3557984]

16. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One 2020;15(3):e0230548 [FREE Full text] [doi: 10.1371/journal.pone.0230548] [Medline: 32191764]

17. Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, Guo Y, et al. Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv Preprint posted online March 3, 2020. [doi: 10.1101/2020.02.27.20028027]

18. Qi X, Jiang Z, Yu Q, Shao C, Zhang H, Yue H, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv Preprint posted online March 3, 2020. [doi: 10.1101/2020.02.29.20029603]

19. Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. SSRN J Preprint posted online February 28, 2020. [doi: 10.2139/ssrn.3543603]

20. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv Preprint posted online February 25, 2020. [doi: 10.1101/2020.02.24.20027268]

21. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. SSRN J Preprint posted online April 6, 2020. [doi: 10.2139/ssrn.3562456]

22. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3]

23. Li L, Huang T, Wang Y, Wang Z, Liang Y, Huang T, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol 2020 Jun;92(6):577-583 [FREE Full text] [doi: 10.1002/jmv.25757] [Medline: 32162702]

24. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). J Gen Intern Med 2020 May;35(5):1545-1549 [FREE Full text] [doi: 10.1007/s11606-020-05762-w] [Medline: 32133578]

25. Thomas L, Reyes EM. Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R. J Stat Soft 2014;61(1):1-23. [doi: 10.18637/jss.v061.c01]

26. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007 Mar 13;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]

27. Heagerty P, Lumley T, Pepe M. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000 Jun;56(2):337-344. [doi: 10.1111/j.0006-341x.2000.00337.x] [Medline: 10877287]

28. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982;69(1):239-241. [doi: 10.1093/biomet/69.1.239]

29. R: A Language and Environment for Statistical Computing. R Core Team. 2020. URL: https://www.r-project.org/

30. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Interac 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]

31. DOCL outcome prediction platform (doctors on the cloud). Digital Health Atlas. URL: https://digitalatlas.who.int/p/00lNbWygey [accessed 2020-05-03]

32. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 2018 Nov;35(11):1763-1774 [FREE Full text] [doi: 10.1007/s12325-018-0805-y] [Medline: 30357570]

33. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. J Infect 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.021] [Medline: 32335169]

34. Bhargava A, Fukushima EA, Levine M, Zhao W, Tanveer F, Szpunar SM, et al. Predictors for severe COVID-19 infection. Clin Infect Dis 2020 May 30:ciaa674 [FREE Full text] [doi: 10.1093/cid/ciaa674] [Medline: 32472676]

35. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 1997 Jan 23;336(4):243-250. [doi: 10.1056/NEJM199701233360402] [Medline: 8995086]

36. Torres OH, Muñoz J, Ruiz D, Ris J, Gich I, Coma E, et al. Outcome predictors of pneumonia in elderly patients: importance of functional assessment. J Am Geriatr Soc 2004 Oct;52(10):1603-1609. [doi: 10.1111/j.1532-5415.2004.52492.x] [Medline: 15450034]

37. Mody L, Sun R, Bradley SF. Assessment of pneumonia in older adults: effect of functional status. J Am Geriatr Soc 2006 Jul;54(7):1062-1067 [FREE Full text] [doi: 10.1111/j.1532-5415.2006.00797.x] [Medline: 16866676]

38. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [doi: 10.1016/S0895-4356(96)00236-3]

39. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995 Dec;48(12):1503-1510. [doi: 10.1016/0895-4356(95)00048-8]

40. Harrell FE. Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer; 2001.

41. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019 Mar 30;38(7):1276-1296 [FREE Full text] [doi: 10.1002/sim.7992] [Medline: 30357870]

42. COVID-19 screening tool. Apple and CDC. URL: https://www.apple.com/covid19/ [accessed 2020-05-04]

43. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

44. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328. [doi: 10.1136/bmj.m1328] [Medline: 32265220]

45. Sperrin M, Grant SW, Peek N. Prediction models for diagnosis and prognosis in Covid-19. BMJ 2020 Apr 14;369:m1464. [doi: 10.1136/bmj.m1464] [Medline: 32291266]

46. Kim ES, Chin BS, Kang CK, Kim NJ, Kang YM, Choi J, Korea National Committee for Clinical Management of COVID-19. Clinical course and outcomes of patients with severe acute respiratory syndrome coronavirus 2 infection: a preliminary report of the first 28 patients from the Korean cohort study on COVID-19. J Korean Med Sci 2020 Apr 06;35(13):e142 [FREE Full text] [doi: 10.3346/jkms.2020.35.e142] [Medline: 32242348]

47. Lean Open Survey on SARS-CoV-2 Infected Patients. 2020. URL: https://leoss.net/ [accessed 2020-05-02]

48. Williams RD, Markus AF, Yang C, Salles TD, Duvall SL, Falconer T, et al. Seek COVER: development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. medRxiv Preprint posted online January 1, 2020. [doi: 10.1101/2020.05.26.20112649]

Abbreviations

AUC: area under the curve
CDC: Centers for Disease Control and Prevention
COVER: COVID-19 Estimated Risk
CURB-65: confusion, urea, respiratory rate, blood pressure, and 65 years or older
EPV: events per variable
HR: hazard ratio
OR: odds ratio
PSSUQ: Post-Study System Usability Questionnaire
VIF: variance inflation factor

XSL•FO
RenderX

Original Paper

Identifying and Ranking Common COVID-19 Symptoms From Tweets in Arabic: Content Analysis

Eisa Alanazi[1], BSc, MSc, PhD; Abdulaziz Alashaikh[2], BEng, MSc, PhD; Sarah Alqurashi[1], BSc; Aued Alanazi[1], BSc

[1]Center of Innovation and Development in Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
[2]University of Jeddah, Jeddah, Saudi Arabia

Corresponding Author:
Eisa Alanazi, BSc, MSc, PhD
Center of Innovation and Development in Artificial Intelligence
Umm Al-Qura University
Prince Sultan Road
Makkah, 11343
Saudi Arabia
Phone: 966 55 601 4171
Email: eaanazi@uqu.edu.sa

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/12/e26446/

Abstract

Background: A substantial amount of COVID-19–related data is generated by Twitter users every day. Self-reports of COVID-19 symptoms on Twitter can reveal a great deal about the disease and its prevalence in the community. In particular, self-reports can be used as a valuable resource to learn more about common symptoms and whether their order of appearance differs among different groups in the community. These data may be used to develop a COVID-19 risk assessment system that is tailored toward a specific group of people.

Objective: The aim of this study was to identify the most common symptoms reported by patients with COVID-19, as well as the order of symptom appearance, by examining tweets in Arabic.

Methods: We searched Twitter posts in Arabic for personal reports of COVID-19 symptoms from March 1 to May 27, 2020. We identified 463 Arabic users who had tweeted about testing positive for COVID-19 and extracted the symptoms they associated with the disease. Furthermore, we asked them directly via personal messaging to rank the appearance of the first 3 symptoms they had experienced immediately before (or after) their COVID-19 diagnosis. Finally, we tracked their Twitter timeline to identify additional symptoms that were mentioned within ±5 days from the day of the first tweet on their COVID-19 diagnosis. In total, 270 COVID-19 self-reports were collected, and symptoms were (at least partially) ranked.

Results: The collected self-reports contained 893 symptoms from 201 (74%) male and 69 (26%) female Twitter users. The majority (n=270, 82%) of the tracked users were living in Saudi Arabia (n=125, 46%) and Kuwait (n=98, 36%). Furthermore, 13% (n=36) of the collected reports were from asymptomatic individuals. Of the 234 users with symptoms, 66% (n=180) provided a chronological order of appearance for at least 3 symptoms. Fever (n=139, 59%), headache (n=101, 43%), and anosmia (n=91, 39%) were the top 3 symptoms mentioned in the self-reports. Additionally, 28% (n=65) reported that their COVID-19 experience started with a fever, 15% (n=34) with a headache, and 12% (n=28) with anosmia. Of the 110 symptomatic cases from Saudi Arabia, the most common 3 symptoms were fever (n=65, 59%), anosmia (n=46, 42%), and headache (n=42, 38%).

Conclusions: This study identified the most common symptoms of COVID-19 from tweets in Arabic. These symptoms can be further analyzed in clinical settings and may be incorporated into a real-time COVID-19 risk estimator.

XSL•FO
RenderX

Introduction

The ongoing COVID-19 pandemic has greatly impacted human health and well-being and has radically enforced a rigorous change in people's lifestyles. In response to this catastrophe, we have witnessed a great effort from diverse research communities to study all aspects of this disease.

In recent years, social networks have become an important source of information where users expose and share ideas, opinions, thoughts, and experiences on a multitude of topics. Several studies have utilized the abundance of information offered by social platforms to conduct nonclinical medical research. For example, Twitter has been a source of data for many health and medical studies, such as surveillance and monitoring of flu and cancer timelines and distribution across the United States [1], analyzing the spread of influenza in the United Arab Emirates based on geotagged tweets in Arabic [2], and the surveillance and monitoring of influenza in the United Arab Emirates based on tweets in Arabic and English [3]. In addition, Twitter data have been utilized in symptom and disease identification in Saudi Arabia [4], and most recently, to examine COVID-19 symptoms as reported on Twitter [5] and to analyze the chronological and geographical distribution of infected tweeters in the United States [6].

The Twitter platform allows researchers to obtain data on items like age, sex, geolocation, etc, along with informative posts, via data mining and analysis techniques; this can potentially result in useful insights about a specific health condition [7]. Extracting common symptoms associated with a disease from publicly available data has the potential to control the spread of the disease and identify users at high risk. It also gives new insights that call for early intervention and control. For example, Figure 1 presents the translation of a tweet (from Saudi Arabia; dating to early May 2020) that explicitly mentions the loss of smell and taste as one distinctive symptom of COVID-19. Interestingly, the official COVID-19 Questionnaire App in Saudi Arabia was updated in late May 2020 to include the sudden loss of smell and taste as one risk indicator of having COVID-19 [8]. Tracking COVID-19 symptoms in real time via public data on Twitter could have shortened the gap.

Figure 1. A patient with COVID-19 tweets about how the loss of smell and taste was the only common symptom across all of their family members. The tweet was anonymized and translated into English.
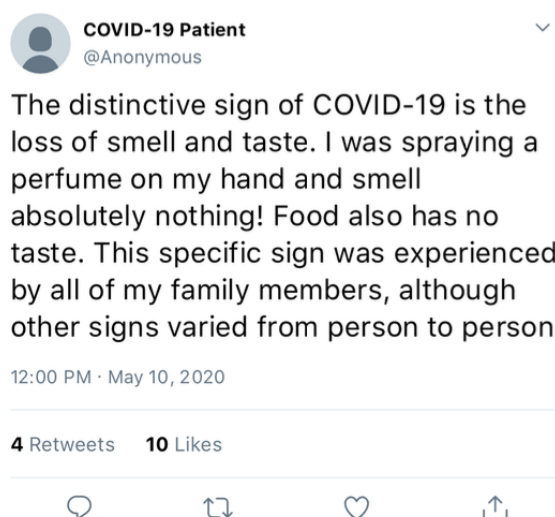
In this paper, we examined COVID-19 symptoms as reported by Arabic tweeters. First, we shuffled tweets in Arabic and searched for tweets with COVID-19 symptoms and collected tweets from users who self-reported a positive diagnosis (via clinical testing). Next, we asked infected users about the first 3 symptoms they had experienced via a voluntary survey sent through a private message.

Methods

Our data collection methodology is outlined in Figure 2. First, we searched Twitter for personal reports of COVID-19 from March 1, 2020, to May 27, 2020, using 2 Arabic keywords [image] and [image], which translate roughly to "I have been diagnosed." Such keywords are likely to filter out reports that were not associated with a formal test result. An initial list of 463 users were collected, and 2 independent freelancers were asked to further read users' timeline and extract symptoms that were explicitly mentioned that were related to COVID-19 and their order of appearance, if mentioned. Additional information such as user gender, date of infection, and country of residence were also collected. We assumed the date of the COVID-19 diagnosis tweet as the date of infection, if no other information was available.

XSL•FO

RenderX

Figure 2. Data collection steps.
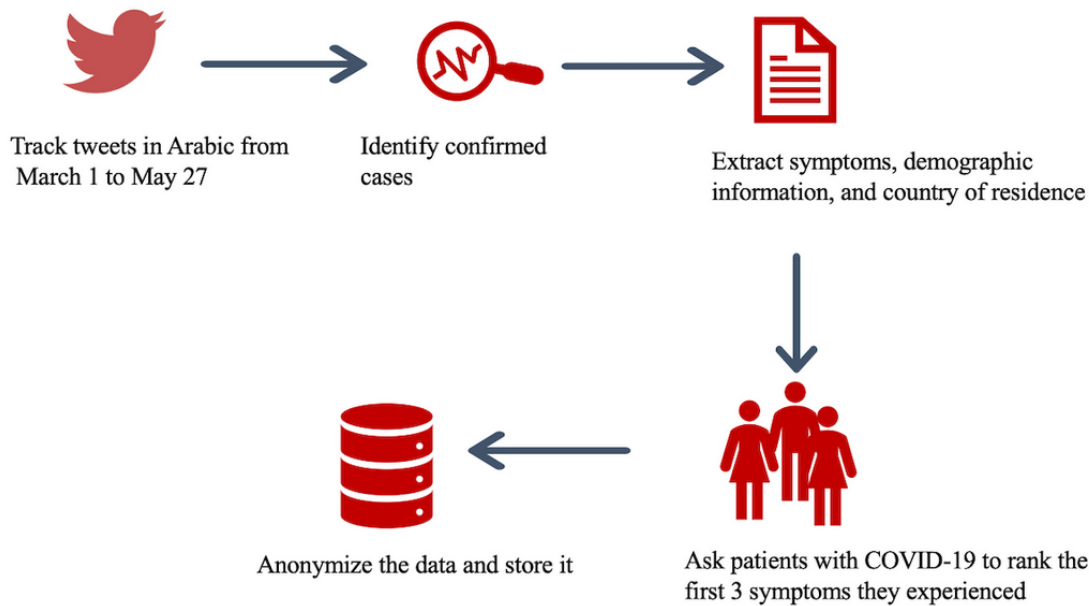
In all, 270 users with COVID-19 were identified, of whom 80 shared their symptoms publicly. To further understand the chronological order of the symptoms, we asked users through Twitter personal messages to rank the first 3 symptoms they experienced right before or after testing positive for COVID-19.

We recorded the symptom ranks (from first to last) based on the received responses and publicly available data on the users' pages. In case no order was given, an implicit order was assumed following the order in which the symptoms were mentioned by the user.

Tracking tweets containing specific keywords is not sufficient enough to obtain an overview of disease dynamics [9]. Many patients detailed their experience while infected; hence, knowing their health condition and sentiments, and tracking useful information, may lead to a better understanding of the disease symptoms. In particular, we found tweets that were posted within ±5 days of infection date to contain valuable information about early symptoms, allowing us to process and rank the symptoms. As an example, Figure 3 highlights 3 tweets by 3 different patients with COVID-19 that indirectly relay symptoms before or after being diagnosed with COVID-19. For simplicity, we set a false date (April 28, 2020) for all 3 tweets using TweetGen [10]. *User..1* tested positive on April 29, 1 day after tweeting their wish to be able to taste food; *User..2* tested positive on May 1, 3 days after complaining about a headache; and *User..3* was tested positive on April 26 and tweeted on April 28 about the loss of smell.

Figure 3. Example of tweets collected within 5 days before or after the user tweeted about having a COVID-19–positive diagnosis.

The examples highlighted in Figure 3 demonstrate that mining Twitter for COVID-19 symptoms requires more than a simple keyword search. In principle, the context of the tweet, as narrated by a user with COVID-19, is also important. Therefore, it is important to examine not only the tweet but also its context.

To build a high-quality database of COVID-19 symptoms based on tweets in Arabic, we relied on a manual extraction of symptoms.

Results

The majority of cases were recorded in May 2020 (n=210, 78%), followed by April (n=39, 14%) and March (n=21, 8%). This surge in May reports is understandable as most countries globally witnessed a substantial increase in the number of confirmed cases. Needless to say, some of the adopted strategies to prevent further spread of the virus (eg, active screening by the Ministry of Health in Saudi Arabia [11]) may have also helped in finding more reports in May compared to other months. We have witnessed this firsthand as some of the

asymptomatic reports were mainly a result of early active screening.

Users from Saudi Arabia, Kuwait, and the United Arab Emirates constituted 85% (n=230) of reports. Nearly half of the reports came from Saudi Arabia (n=125, 46%), which is not surprising, since it is one of the top countries on Twitter with more than 15 million users [12]. Other countries (Egypt, Iraq, Bahrain, Qatar, United Kingdom, United States, Belgium, and Germany) constituted the remaining 15% (n=40).

We collected 893 symptoms from 270 self-reports (as shown in Table 1). The daily number of collected tweets is highlighted in Figure 4.

Table 1. Number of symptoms experienced by tweeters (N=270).

Symptom count	Number of reports, n (%)
0	36 (13)
1	19 (7)
2	35 (13)
3	65 (24)
4	50 (19)
5	35 (13)
6	11 (4)
7	8 (3)
8	5 (2)
9	3 (1)
10	3 (1)

Figure 4. Number of daily collected reports from Twitter (March to May 2020).
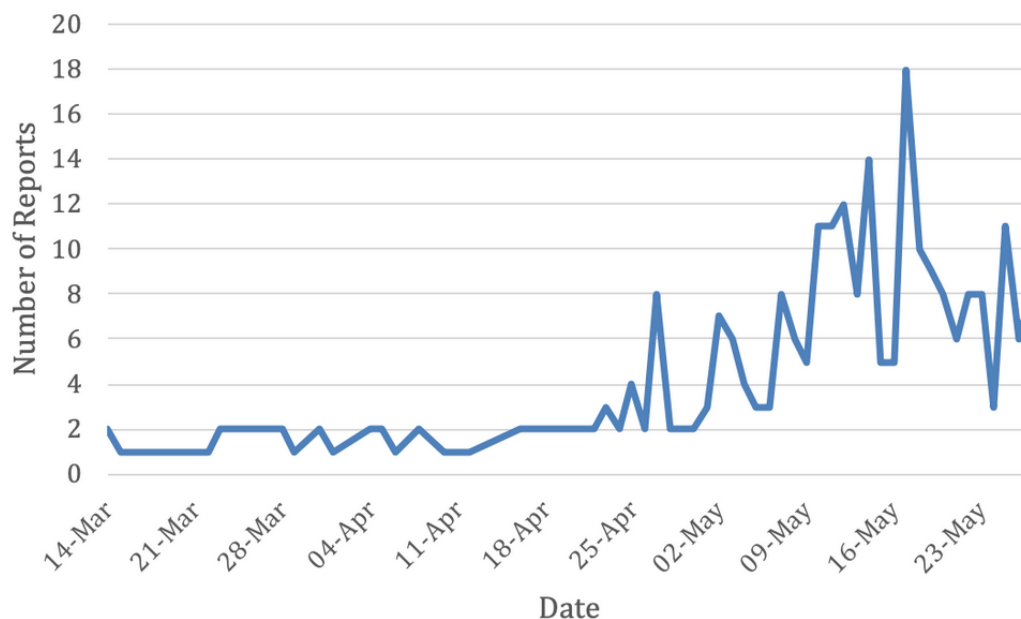
Table 1 indicates that most tweeters experienced between 2 to 5 symptoms, whereas 13% (n=36) of the reported cases were asymptomatic. Table 2 lists the frequency of each symptom ordered from the most prevalent to the least. Only fever was experienced by nearly 60% (n=139) of users. The frequency of

symptoms appears to be consistent across male and female patients (correlation coefficient=0.966). Further, Table 3 lists the top 8 symptoms in the order of their appearance (ie, first, second, and third rank); this refers to the 8 most common symptoms that were ranked as the first, second, or third

symptom to occur in the collected self-reports. Fever and headache were commonly the first reported symptoms. The top 4 symptoms that coincided with fever were headache (n=33, 23.7%), cough (n=20, 14.4%), anosmia (n=19, 13.7%), and ageusia (n=17, 12.2%). Other symptoms occurred at a relatively lower frequency with fever. Table 4 lists the top 8 common symptoms for Saudi Arabia and Kuwait, which accounts for 81.2% (n=190) of the reports. The symptoms had a correlation coefficient of 0.835 between the 2 countries.

Table 2. Most common symptoms reported by users.

Symptom	All users (n=234), n (%)	Male (n=171), n (%)	Female (n=63), n (%)
Fever	139 (59)	98 (57)	41 (65)
Headache	101 (43)	68 (40)	33 (52)
Anosmia	91 (39)	63 (37)	28 (44)
Ageusia	72 (31)	51 (30)	21 (33)
Fatigue	68 (29)	54 (32)	14 (22)
Cough	62 (26)	48 (28)	14 (22)
Sore throat	42 (18)	30 (18)	12 (19)
Dyspnea	33 (14)	26 (15)	7 (11)
Diarrhea	27 (12)	22 (13)	5 (8)
Runny nose	23 (10)	17 (10)	6 (9)
Arthralgia	16 (7)	10 (6)	6 (9)
Chest pain	15 (6)	13 (8)	2 (3)
Back pain	14 (6)	11 (6)	3 (5)
Anorexia	14 (6)	11 (6)	3 (5)
Body ache	12 (5)	8 (5)	4 (6)
Nausea	12 (5)	8 (5)	4 (6)
Osteodynia	11 (5)	8 (5)	3 (5)
Dry throat	9 (4)	6 (3)	3 (5)
Myalgia	9 (4)	7 (4)	2 (3)
Dizziness	8 (3)	6 (3)	2 (3)
Chills	7 (3)	5 (3)	2 (3)
Nasal congestion	7 (3)	4 (2)	1 (2)
Sinusitis	7 (3)	3 (2)	4 (6)

Table 3. The top 8 symptoms, with a first, second, and third rank, as reported by users.

Number	First	Second	Third
1	Fever	Fever	Fever
2	Headache	Headache	Headache
3	Anosmia	Fatigue	Anosmia
4	Fatigue	Cough	Ageusia
5	Cough	Ageusia	Fatigue
6	Sore throat	Anosmia	Cough
7	Runny nose	Sore throat	Anorexia
8	Diarrhea	Arthralgia	Dyspnea

XSL•FO
RenderX

Table 4. The top 8 common symptoms for Saudi Arabia and Kuwait.

Symptom	Saudi Arabia (n=110), n (%)	Kuwait (n=80), n (%)
Fever	65 (59)	45 (56)
Headache	42 (38)	38 (48)
Anosmia	46 (42)	21 (26)
Ageusia	36 (37)	19 (24)
Fatigue	31 (28)	19 (24)
Cough	21 (19)	19 (24)
Sore throat	22 (20)	11 (14)
Dyspnea	14 (13)	11 (14)

Finally, we compared the symptom prevalence of our study to the one provided by Sarker et al [5], in order to assess similarities and differences in COVID-19 symptoms experienced by different populations. As seen in Table 5 and Figure 5, our findings complement those of Sarker et al [5] (correlation coefficient=0.72).

Table 5. Comparison of common symptoms found in this study and in Sarker et al [5].

Symptom	Our study (n=234), n (%)	Sarker et al (n=171), n (%)
Fever	139 (59)	113 (66)
Headache	101 (43)	64 (37)
Anosmia	91 (39)	49 (29)
Ageusia	72 (31)	48 (28)
Fatigue	68 (29)	72 (42)
Cough	62 (26)	99 (58)
Sore throat	42 (18)	41 (24)
Dyspnea	33 (14)	62 (36)
Diarrhea	27 (12)	15 (9)
Runny nose	23 (10)	16 (9)
Arthralgia	16 (7)	2 (1)
Chest pain	15 (6)	39 (23)
Back pain	14 (6)	—[a]
Anorexia	14 (6)	23 (14)
Body ache	12 (5)	73 (43)
Nausea	12 (5)	19 (13)
Osteodynia	11 (5)	—
Dry throat	9 (4)	—
Myalgia	9 (4)	10 (6)
Dizziness	8 (3)	15 (9)
Chills	7 (3)	43 (25)
Nasal congestion	7 (3)	—
Sinusitis	7 (3)	7 (4)

[a]Not applicable.

Figure 5. A comparison between symptom prevalence in our study and Sarker et al [5] (correlation coefficient=0.72).
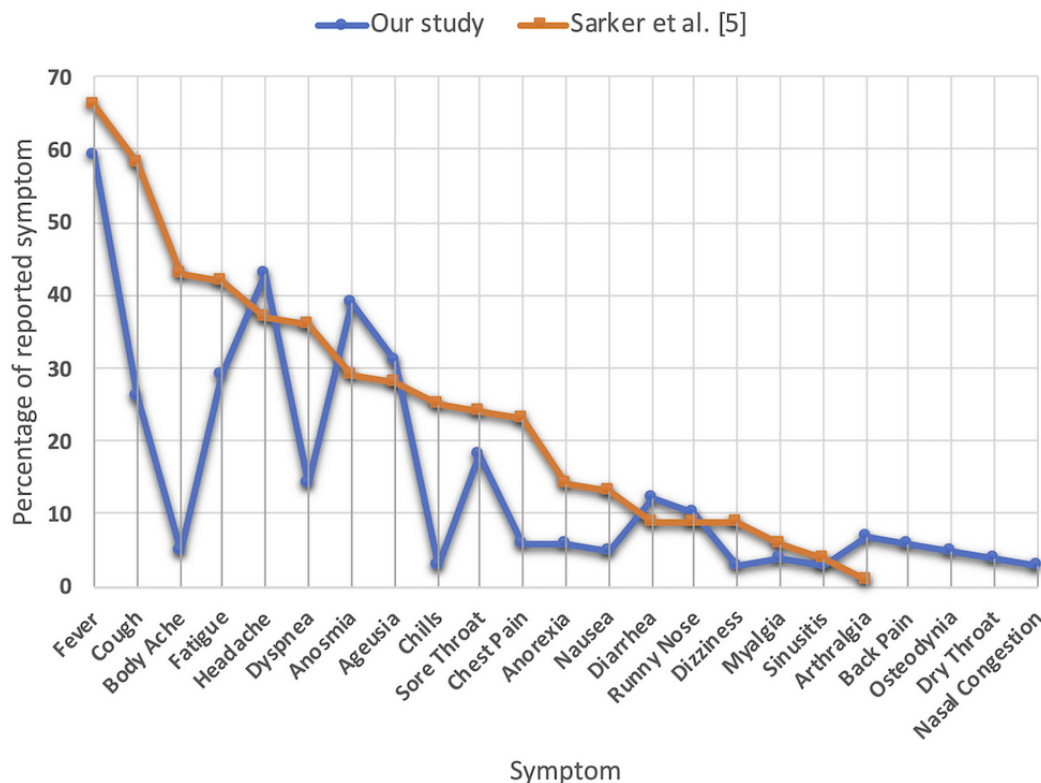
Discussion

Principal Findings

This work identified common COVID-19 symptoms from Arabic personal reports on Twitter. These findings complement the results of other recent studies [5,6,9] that focused on tweets in English or specific demographic groups. This study was carried in a way to report not only the symptoms but their timeline as narrated by users. Social networks have become the de facto communication channel for a large number of people. Many individuals worldwide write, interact, or even just browse social network content countless times a day. Social networks have the property of being continuously updated with new information provided by other global citizens. As such, it is crucial to monitor their content to identify health issues [13,14]. One potential benefit of analyzing social networks is understanding COVID-19 symptoms and identifying people at high risk [7].

Anosmia being one of the top 3 reported symptoms, mentioned in 39% of reports, was a surprising result in our study. Several tweeters complained about the longevity of anosmia. Our sample size is still relatively small to make any sound judgment in this regard. However, recent clinical studies have reported finding anosmia in 35.7% of mild cases of COVID-19, which is relatively close to our estimation from the tweets examined in this study [15]. In fact, the number of self-reports reflects the testing capacity of different countries. As of June 9, 2020, Saudi Arabia had completed almost 1 million tests, and Kuwait had carried out more than 350,000 tests [16].

It is worth noting that some users experienced weight loss due to COVID-19; one user claimed losing 20 kg due to the disease.

Another interesting observation is that several users experienced what they described as a short-term mild fever for a couple of hours only. Quitting smoking was a positive outcome of COVID-19, per one user's tweet. We were surprised by some users in early April claiming to be positive for COVID-19, which later turned out to be an April Fool's Day prank. These findings prompt further study into how different communities react to a pandemic and how it affects their lives.

Limitations and Future Works

Several limitations need to be acknowledged. Self-reports from Egypt, the largest Arabic country with almost 100 million people, were inadequately represented in this study. This could be attributed to factors such as Egypt's preference for other social media platforms (eg, Facebook), as well as differing dialects and use of local idioms.

Our study tracked 2 widely used keywords to identify Arabic patients with COVID-19 on Twitter, followed by a manual extraction of symptoms. More complex keywords could reveal additional interesting patterns about symptoms. Furthermore, we used Modern Standard Arabic (MSA) keywords to obtain a general view of Twitter content in Arabic. It is, however, well noted in the literature that many Arabic users write in their own local dialect on social media. Hence, it is helpful to consider not only keywords in the MSA form but also keywords that are tailored toward different Arabic dialects to better capture tweets on COVID-19 symptoms written in Arabic. This may explain why Egypt was underrepresented in this study. Therefore, a multidialect COVID-19 Arabic dictionary and an natural language processing–based algorithm to detect and analyze tweets in Arabic need to be developed; establishing a comprehensive medical dictionary for different local Arabic

dialects is an important line of research during the coronavirus pandemic [17].

We have extracted symptoms from users who likely underwent a screening test and, hence, tweeted based on its result; however, we do not have confirmation of testing. In this study, we have not used other COVID-19 sources; specifically, studying personal reports in Arabic from both Facebook and Twitter would have enhanced study results.

The noticeable increase in May reports compared to other months demonstrates the importance of developing a real-time surveillance system based on the symptoms reported in Twitter posts in Arabic. It also suggests further studies of information sharing behaviors in different communities and across different demographic groups (ie, users grouped by age, gender, geolocation, etc) are needed [18].

One interesting observation from our analysis is related to gender distribution. Approximately 25% of the collected reports came from female users. This could be due to several reasons. One reason could be the presence of more male Arabic patients with COVID-19 than female ones; however, we are not aware of any reliable source to support this claim. Nevertheless, in Saudi Arabia, cases reported by males consistently outnumbered those reported by females in April and May 2020 [19]. Further

insights and studies are needed to investigate the gender differences in information sharing behaviors and analyze whether there is any notable difference in how male and female Arabic users disclose health information on social media.

Privacy is one of the key issues that needs to be addressed before utilizing social media for public health surveillance. Apart from each network's privacy policy, there exists no global concensus on what to disclose when collecting health information from social media networks. Some attempts in the literature have suggested best practices to follow when collecting health information from Twitter [20]. Such practices include, among other things, avoiding quoting directly from users' tweets and mentioning users' IDs. Moreover, some social media sites have updated their privacy policy to further control content redistribution. For instance, Twitter's updated policy permits redistribution of only the tweets' ID and not their content verbatim to third parties [21].

Conclusion

This study identified the most common self-reported COVID-19 symptoms from tweets in Arabic. Our findings demonstrated that fever, headache, and anosmia are the 3 most common symptoms experienced by users, and we presented symptom prevalence for two of the largest clusters found in our tweets database (Saudi Arabia and Kuwait).

Authors' Contributions

EA and A Alashaikh designed the study and wrote the manuscript. SA developed the social network analysis methodology and collected related tweets using Twitter API. A Alanazi extracted and translated the symptoms collected from personal reports to their scientific names. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Lee K, Agrawal A, Choudhary A. Real-time disease surveillance using Twitter data: demonstration on flu and cancer. 2013 Presented at: KDD '13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining; August 2013; Chicago, Illinois, USA p. 1474-1477. [doi: 10.1145/2487575.2487709]

2. Alkouz B, Al Aghbari Z. Analysis and prediction of influenza in the UAE based on Arabic tweets. 2018 Presented at: IEEE 3rd International Conference on Big Data Analysis (ICBDA); 9-12 March 2018; Shanghai, China p. 61-66. [doi: 10.1109/icbda.2018.8367652]

3. Alkouz B, Aghbari ZA, Abawajy JH. Tweetluenza: Predicting flu trends from twitter data. Big Data Min. Anal 2019 Dec;2(4):273-287. [doi: 10.26599/bdma.2019.9020012]

4. Alotaibi S, Mehmood R, Katib I, Rana O, Albeshri A. Sehaa: A Big Data Analytics Tool for Healthcare Symptoms and Diseases Detection Using Twitter, Apache Spark, and Machine Learning. Applied Sciences 2020 Feb 19;10(4):1398 [FREE Full text] [doi: 10.3390/app10041398]

5. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang YC. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. J Am Med Inform Assoc 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: 10.1093/jamia/ocaa116] [Medline: 32620975]

6. Klein A, Magge A, O'Connor K, Cai H, Weissenbacher D, Gonzalez-Hernandez G. A Chronological and Geographical Analysis of Personal Reports of COVID-19 on Twitter. medRxiv 2020 Apr 22 [FREE Full text] [doi: 10.1101/2020.04.19.20069948] [Medline: 32511608]

XSL•FO
RenderX

7.   Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1). [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]

8.   Ministry of Health Saudi Arabia. The Central Appointment System (Mawid). Ministry of Health E-Services. URL: https://www.moh.gov.sa/en/eServices/Pages/cassystem.aspx [accessed 2020-05-27]

9.   Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2) [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]

10. Tweetgen. 2020. URL: https://www.tweetgen.com/ [accessed 2020-06-02]

11. Nugali N. Saudi Arabia's active mass testing contains COVID-19 spread. Arab News. 2020 Apr 21. URL: https://www.arabnews.com/node/1662856/saudi-arabia [accessed 2020-05-25]

12. Leading countries based on number of Twitter users as of April 2020. Statista. 2020. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2020-05-28]

13. Paul M, Sarker A, Brownstein J, Nikfarjam A, Scotch M, Smith K, et al. Social media mining for public health monitoring and surveillance. Biocomputing 2016 2016:468-479 [FREE Full text] [doi: 10.1142/9789814749411_0043]

14. Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. PLoS One 2013 Apr 24;8(4) [FREE Full text] [doi: 10.1371/journal.pone.0061373] [Medline: 23637820]

15. Levinson R, Elbaz M, Ben-Ami R, Shasha D, Levinson T, Choshen G, et al. Time course of anosmia and dysgeusia in patients with mild SARS-CoV-2 infection. medRxiv. Preprint posted online April 14, 2020 [FREE Full text] [doi: 10.1080/23744235.2020.1772992] [Medline: 32552475]

16. Worldometer. COVID-19 Pandemic. 2020. URL: https://www.worldometers.info/coronavirus/ [accessed 2020-06-09]

17. Guo X, Li J. A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. 2019 Presented at: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS); 22-25 Oct 2019; Granada, Spain. [doi: 10.1109/snams.2019.8931720]

18. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug 19;15(8) [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]

19. Saudi Arabic Coronavirus disease (COVID-19) situation. KAPSARC. 2020. URL: https://datasource.kapsarc.org/explore/dataset/saudi-arabia-coronavirus-disease-covid-19-situation/information [accessed 2020-07-22]

20. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014 Dec 22;16(12) [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]

21. Twitter. Twitter Developer Policy. URL: https://developer.twitter.com/en/developer-terms/policy [accessed 2020-04-15]

Abbreviations

MSA: Modern Standard Arabic

XSL•FO
RenderX

Original Paper

A Patient Self-Checkup App for COVID-19: Development and Usage Pattern Analysis

JoonNyung Heo[1*], MD; MinDong Sung[2*], MD; Sangchul Yoon[3,4*], MD, PhD; Jinkyu Jang[5], PhD; Wonwoo Lee[6], MD; Deokjae Han[7], MD; Hyung-Jun Kim[7], MD; Han-Kyeol Kim[6], MD; Ji Hyuk Han[8], MD, MPH; Woong Seog[1], MD; Beomman Ha[1], MD; Yu Rang Park[2], PhD

[1]Armed Forces Medical Command, Seongnam, Republic of Korea

[2]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Medical Humanities and Social Sciences, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Ophthalmology, Yonsei University College of Medicine, Seoul, Republic of Korea

[5]Companoid Labs, Yonsei University, Seoul, Republic of Korea

[6]Department of Neurology, Yonsei University College of Medicine, Seoul, Republic of Korea

[7]Department of Internal Medicine, The Armed Forces Capitol Hospital, Seongnam, Republic of Korea

[8]Department of Otorhinolaryngology, The Armed Forces Capitol Hospital, Seongnam, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 222282493
Email: yurangpark@yuhs.ac

Abstract

Background: Clear guidelines for a patient with suspected COVID-19 infection are unavailable. Many countries rely on assessments through a national hotline or telecommunications, but this only adds to the burden of an already overwhelmed health care system. In this study, we developed an algorithm and a web application to help patients get screened.

Objective: This study aims to aid the general public by developing a web-based application that helps patients decide when to seek medical care during a novel disease outbreak.

Methods: The algorithm was developed via consultations with 6 physicians who directly screened, diagnosed, and/or treated patients with COVID-19. The algorithm mainly focused on when to test a patient in order to allocate limited resources more efficiently. The application was designed to be mobile-friendly and deployed on the web. We collected the application usage pattern data from March 1 to March 27, 2020. We evaluated the association between the usage pattern and the numbers of COVID-19 confirmed, screened, and mortality cases by access location and digital literacy by age group.

Results: The algorithm used epidemiological factors, presence of fever, and other symptoms. In total, 83,460 users accessed the application 105,508 times. Despite the lack of advertisement, almost half of the users accessed the application from outside of Korea. Even though the digital literacy of the 60+ years age group is half of that of individuals in their 50s, the number of users in both groups was similar for our application.

Conclusions: We developed an expert-opinion–based algorithm and web-based application for screening patients. This innovation can be helpful in circumstances where information on a novel disease is insufficient and may facilitate efficient medical resource allocation.

XSL•FO
RenderX

Introduction

On March 11, 2020, the World Health Organization officially characterized COVID-19, a disease caused by the novel coronavirus SARS-CoV-2, as a pandemic [1]. As general preventive guidelines, hygiene control and social distancing were recommended, but in a situation where one shows signs of infection, there are no clear guidelines for patients on how to respond. Many countries, including South Korea, rely on individual assessments and advice provided by a health care worker. This could be carried out by calling a national hotline or one's local clinic, or by means of telecommunication that assesses and advises people on what to do in the event of a suspected infection [2]. This often results in overwhelmed call centers with long waiting times and untimely response to emergency patients.

Viral pandemics such as COVID-19 can place extraordinary and sustained demands on public health systems and essential community service providers [3]. Patient triage is essential in such situations to efficiently allocate resources [4,5]. In addition, patients need to be advised on the adequate care measures to be taken not only for themselves but also for others around them. However, for novel infections like COVID-19, it is difficult for both the general public and less experienced medical staff to accurately make decisions.

Traditionally, for medical advice to be made public, many trials have to repetitively prove that the advice is sound and without any unforeseen side effects. This is, in many ways, appropriate, considering the consequences of erroneous recommendations. However, in the current situation, patients who stand a chance of survival with professional care may still die due to a lack of clinical resources. Thus, a well-timed, imperfect solution may be more beneficial than belated, flawless advice. In addition, this information can easily be confusing to the general public, and false interpretation may cause harm to users. Providing medical knowledge in a friendly and clear manner through an application can be both satisfactory and safe for users. Providing screening algorithms based on experiences from actively involved physicians can help improve the efficiency of the diagnosis process. This is made possible by both accurately selecting the patients to be tested and alleviating the workload imposed on health care workers.

In this study, we developed an algorithm and a web application for use by the general public to help them decide whether or not to seek professional care in cases where a COVID-19 infection is suspected.

Methods

Algorithm Development

During the initial phases of the COVID-19 outbreak in Korea, the government and the Korean Center for Disease Control and Prevention (KCDC) demonstrated expansive and extensive efforts to isolate infected people and trace and quarantine their contacts. The KCDC guidelines stated that two groups of patients be tested: suspected cases and patients under investigation [6]. A person exhibiting fever or respiratory symptoms within 14 days of contact with a confirmed COVID-19 case was defined as a suspected case. Patients under investigation were defined as those who traveled overseas or those with an epidemiological link to a domestic COVID-19 cluster with fever or respiratory symptoms. In addition, a person suspected to have COVID-19 according to a physician's opinion was also classified as a patient under investigation and was subject to COVID-19 testing. Many patients were tested based on their physician's opinion, and no clear guidelines existed on features to be considered.

Our algorithm was built with the aim of helping a patient decide when they should consult a doctor for COVID-19 testing. It was developed through the consultation of a group of physicians (n=6) directly involved in the process of screening, diagnosis, and/or treatment of patients with COVID-19. The KCDC guidelines were also reviewed for any information on patient selection for COVID-19 testing [6].

The physician group comprised 1 pulmonologist and 1 infectious disease specialist directly involved in the care of confirmed cases; 1 pulmonologist and 1 otorhinolaryngologist involved in patient screening at the hospital; and 2 neurologists involved in patient screening at public health centers. Based on experience and a literature review, clinical variables that should be considered when screening for patients with COVID-19 were selected, and the decision process of selecting patients to be tested for COVID-19 was discussed.

The final algorithm of COVID-19 CheckUp is shown in Figure 1. Upon reviewing the KCDC guidelines [6] and several papers in the literature [7-9], the patients who were required to report to the KCDC and undergo a COVID-19 test were clearly defined. KCDC guidelines state that patients who show any symptoms of upper respiratory infection or documented fever combined with direct contact with an individual with confirmed COVID-19 or have visited a well-known outbreak area should report to the KCDC and take a COVID-19 confirmatory test immediately.

Figure 1. Algorithm for the COVID-19 CheckUp app. BT: body temperature.
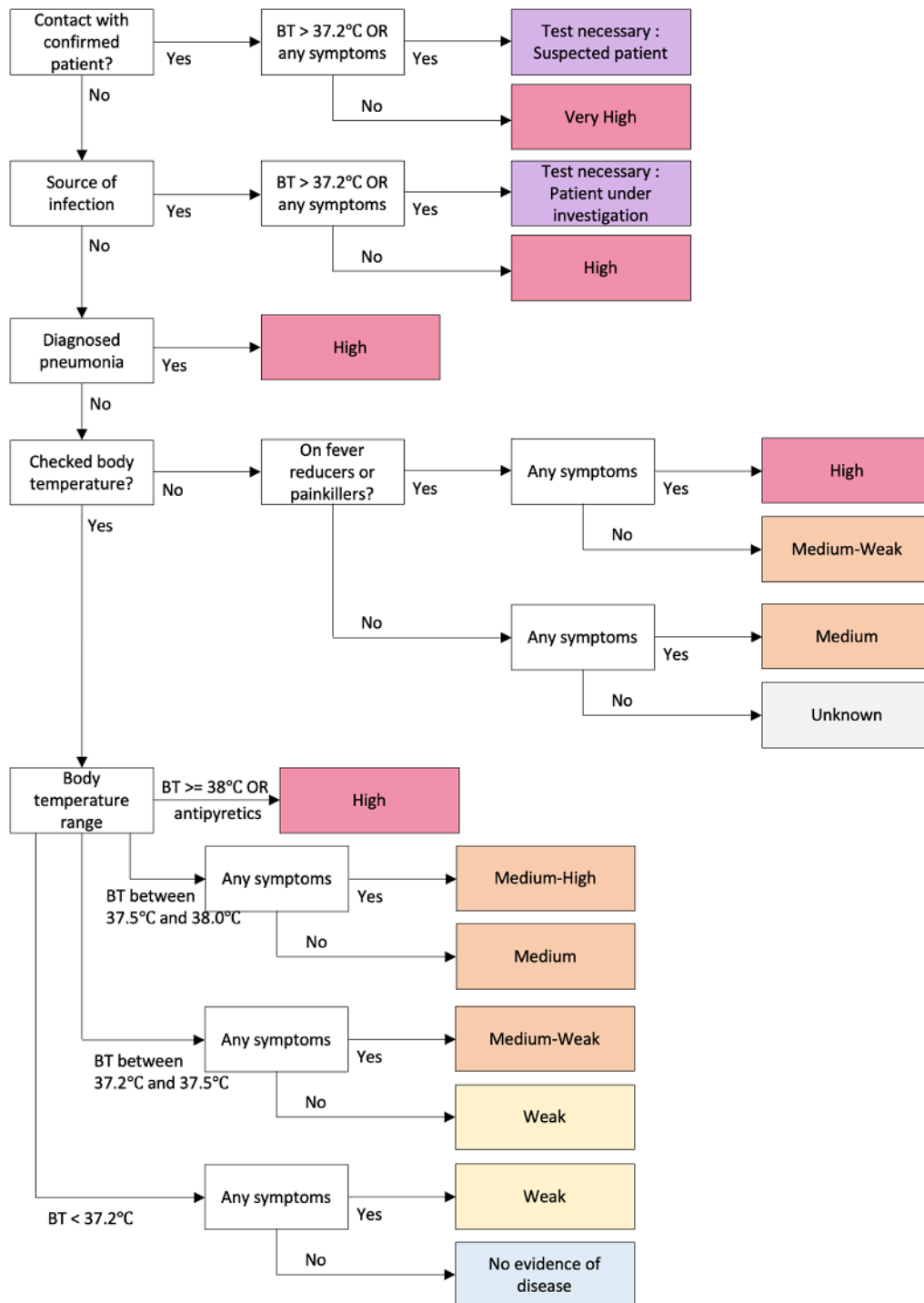
The first part of the algorithm was designed to meet the KCDC guidelines. The user is asked if there was any contact with a confirmed patient or any other source of infection. If the user answered "yes" to these questions and presented any relevant symptoms or fever, they are instructed to report to the KCDC and take a COVID-19 test. A user who had contact a confirmed COVID-19 case is still classified as highly likely to be needing a test even if the patient presents no symptoms and has a normal body temperature.

Body temperature was classified into four categories: high fever (higher than 38 °C), fever (higher than 37.5 °C but lower than 38 °C), mild fever (higher than 37.2 °C but lower than 37.5 °C), and normal (lower than 37.2 °C). Body temperature was chosen as the most important variable for the process of screening as it was one of the most objective variables the user could provide; further, the pathogens in patients with high fever were more likely to be viral. In addition, patients using antipyretics, including nonsteroidal anti-inflammatory drugs (NSAIDs), were

treated like patients with high fever. This was done to reflect cases where fever and symptoms of upper respiratory infection were masked by the chronic use of antipyretics; additionally, patients who took these medications were usually at a higher risk because of their chronic comorbidities. The detailed classification of temperature into four categories (high fever, fever, mild fever, and normal) was used to consider the numerous reports of patients with COVID-19 presenting with only mild fever [10], and at the same time differentiate patients with high fever, who should be considered more gravely.

Application

COVID-19 CheckUp was developed using the Flutter software development kit [11]. It was deployed as a web application [12]

for better accessibility and maintenance. The application has been designed to be mobile-friendly and interactive. HTML5 local storage was used to store users' data, which ensures that usage data are never transferred to a server. This structure ensures that the privacy of users' data can be protected and at the same time eliminates the need for a back-end architecture. To better accommodate international usage, the application has been translated into 5 languages: Korean, English, French, Spanish, and Vietnamese. Screenshots of the COVID-19 CheckUp app are shown in Figure 2. When using the app, the patient is asked a maximum of 7 simple questions based on our algorithm.

Figure 2. Screenshots of the COVID-19 CheckUp app.

Usage Patterns

To gain insight on application usage, Google Analytics [13] was embedded and usage data were collected from March 1 to March 27, 2020. Google Analytics API [14] was used to further help extract data from Google Analytics. The extracted data include frequency of access and location of the user. The data on COVID-19 cases and deaths were extracted from Johns Hopkins University's COVID-19 data repository on GitHub [15]. The usage numbers were compared by subcontinent and city. Screened cases for each city were also included for analysis. Each local autonomous body in Korea released the number of COVID-19 tests that were conducted. For Indonesia, because data on the number of assessments were not available, the number of screened cases was defined as the sum of patients classified as "under monitoring," "under supervision," and "confirmed." Moreover, usage by age group was compared to their corresponding digital literacy and COVID-19 fatality by age (sourced from the KCDC). Digital literacy by age was

obtained from a report on the digital divide in 2019 released by the Ministry of Science and ICT and the National Information Society Agency in Korea [16].

All the statistical analyses were performed using R 3.6.0 (The R Foundation).

Results

In total, 83,640 users accessed the application 105,508 times from 141 countries. In Eastern Asia, where Korea is located, 43,648 users accessed the application 57,030 times. Table 1 reports the number of application users by subcontinent compared with the number of COVID-19 confirmed cases and deaths. Although the application had not been introduced or advertised outside of Korea, the website received heavy traffic from other subcontinents in the following order: Southeast Asia, North America, and Western Europe (Multimedia Appendix 1).

Table 1. The number of users, new users, sessions, COVID-19 cases, and deaths by subcontinent.

Subcontinent	Users (n=83,640), n	Sessions (n=105,508), n	Cases (n=7,584,046), n	Deaths (n=289,731), n
Eastern Asia	43,648	57,030	4,103,263	135,335
Southeast Asia	31,117	38,033	52,604	1083
North America	5767	6653	499,451	7112
Western Europe	734	881	817,082	16,536
Australasia	575	661	21,042	125
Northern Europe	400	495	200,750	4571
Western Asia	301	331	75,591	875
South America	263	356	49,203	748
Southern Europe	251	315	1,271,012	98,153
Eastern Europe	136	164	44,315	400
Central America	83	100	9037	114
Southern Asia	78	102	384,601	23,763
Central Asia	66	84	1516	6
Northern Africa	64	77	10,085	477
Eastern Africa	53	94	1509	20
Sub-Saharan Africa	25	33	3558	73
Southern Africa	21	30	5122	1
Caribbean	20	25	4001	85
Unidentified	16	17	—[a]	—
Micronesia	15	19	—	—
Middle Africa	7	8	1073	25

[a]Not applicable.

[Table 2](#) shows the list of the top 10 cities with the highest user ratios (user count per 100,000 population). The majority of users were from Korea (n=43,210, 51.64%); almost half (n=40,907, 48.36%) were from outside of Korea.

In Korea, the digital literacy of users in their 50s and 60s is 73.60% and 35.70%, respectively. However, application usage was similar between the two groups at 5.5% each. Even though the digital literacy of users in their 60s was half that of users in their 50s, the user count was similar for our application ([Table 3](#)).

Table 2. Top 10 cities with the highest user ratios.

City	Country	Users[a], n	Population, N	User ratio[b]	Confirmed, n	Confirmed ratio[b]	Screened[c], n	Screened ratio[b]
Seoul	South Korea	22,542	10,010,983	225.17	360	3.60	65,952	658.80
Busan	South Korea	4768	3,459,840	137.81	112	3.24	__d	—
Jakarta	Indonesia	13,646	10,504,100	129.91	515	4.90	3704	35.26
Depok	Indonesia	2866	2,727,209	105.09	10	0.37	634	23.25
Bandung	Indonesia	2093	2,580,191	81.12	3	0.12	221	8.57
Daegu	South Korea	1805	2,468,222	73.13	6482	262.62	—	—
Surabaya	Indonesia	2110	2,944,403	71.66	31	1.05	217	7.37
Daejeon	South Korea	921	1,493,979	61.65	31	2.07	8201	548.94
Incheon	South Korea	1293	3,029,285	42.68	47	1.55	15,219	502.40
New York	United States	534	8,398,748	6.36	20,011	238.26	—	—

[a]Users from March 1 to 27, 2020.

[b]Ratio defined as count per 100,000 people.

[c]In Korea, screened cases indicate the patients who underwent COVID-19 testing, but in Indonesia, the number of screened cases were defined as the sum of patients classified as under monitoring, under supervision, and confirmed.

[d]Not available.

Table 3. Application usage, digital literacy, and mortality, compared by age group, in Korea.

Variable	Age group					
	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	≥70 years
Application usage[a], n (%)	23,001 (27.50)	28,019 (33.50)	12,964 (15.50)	10,455 (12.50)	4600 (5.50)	4600 (5.50)
Digital literacy[b] (%)	112.30	123.00	121.70	112.70	73.60	35.70
Fatality cases, n (%)[c]	0 (0)	1 (0.10)	1 (0.08)	10 (0.56)	21 (1.75)	111 (10.43)

[a]Estimated users from age group percentage and total number of users.

[b]Digital literacy is expressed as a relative score to the average literacy of the Korean public (score=100). A group score >100 indicates that the digital literacy of that group is higher than that of the general public.

[c]The COVID-19 fatality rate (%) is defined as an occurrence of death from a confirmed case of COVID-19.

Discussion

Principal Findings

This study explains the development and deployment of a globally used COVID-19 symptom checkup application. The main focus of this solution was to provide information about the coronavirus disease, including which symptoms should be taken seriously and under what circumstances COVID-19 should be suspected. Many measures were taken to "flatten the curve," but there were not enough medical resources to smoothly handle the overwhelming spread of the virus [17]. Inevitably, there was mass confusion and distress among the public, which in turn worsened the shortage of medical resources. Our application aimed to break this cycle with, however imperfect, timely knowledge of the unknown virus. This study describes the development of an application to screen for a novel emerging infectious disease that can be deployed worldwide. The usage pattern of the application was not proportional to digital literacy. The application was quickly adapted for use in the military and the government.

Medical resources may not always be sufficient, especially in a pandemic situation. Traditionally, consultation through telecommunication, such as a national hotline, was used to triage patients. This method may be more thorough than a simple decision support system. However, considering how the spread of the virus is overwhelming the entire health care industry, using ICT (information and communications technology) to alleviate the workload of health care workers can be beneficial.

Upon building the algorithm for the application, all the participating physicians easily concluded that there cannot be a decisive method to screen patients for COVID-19 infection. This is mainly because many patients are asymptomatic, with some studies reporting that up to 75% of patients can show no symptoms [18-21]. Hence, the algorithm presents a "no evidence" rather than "testing not needed" result when there is no evidence to indicate COVID-19 infection. A total of 10 levels of risk are presented to the user, ranging from "test necessary" and "very high" to "unknown" and "no evidence." The fine division of the risk level is to inform the user of the factors that

should be considered, and which factors are more important than others, for testing.

Web-based applications are easily accessible because they can be used with a typical preinstalled web browser, rather than having to install a separate app. Moreover, even though our application was developed for a mobile interface, it can be used on any devices such as desktops that connect to the internet. This wide range of supported devices may be beneficial for environments where a mobile device is unavailable.

Without any advertising abroad, the fast adoption of our application overseas reflects the need for such a solution amidst the general public. The application was translated into a total of 5 languages, increasing accessibility for global users. Referrer data acquired using the Google Analytics tool showed that 20.95% of user acquisition was from Twitter, which may account for the worldwide usage of the application. Additionally, the application was registered in the World Health Organization Digital Health Atlas [22].

There are concerns regarding ICT accessibility (eg, too complicated to use or unfamiliar to an older audience). However, our study shows that the user distribution is unexpectedly higher in the older age groups with respect to the digital literacy rate of older adults [16]. Considering that disease fatality increases among elderly patients [23], users' needs seem to outweigh the discomfort of new technologies. The application is constructed as a series of pages with only one question per page. This simple question-and-answer type interface may be more familiar to the elderly user group. However, additional research is needed to confirm the hypothesis.

Comparison With Prior Works

There are other applications that help users to screen themselves. Apple Inc, in association with the Centers for Disease Control and Prevention, released the web-based COVID-19 Screening Tool [24]. In addition to physicians, the app also focused on patients, and it was found to be useful in the United States. Spector and colleagues developed the COVID-19 Symptom Tracker app in the United Kingdom, with a partnership between researchers at King's College London and the health-related data science company ZOE [25]. Judson et al [26] introduced a self-triage and a self-scheduling tool based on well-designed algorithms. Integrated into the electronic medical record (EMR), it enables physicians to easily use the application and possibly enable automatic data collection of user-provided data.

EMR-tethered solutions are effective for local users but not for use outside of the region where the associated EMR system is not available. In Korea, the Self-Check Mobile App has been developed by the Ministry of Health & Welfare to monitor symptoms in people for 14 days who arrived in Korea from abroad [27]. We provide a list of similar applications in Multimedia Appendix 2.

The application described in this paper, the COVID-19 CheckUp app, is officially in use in military hospitals and throughout the military service line. Use of the app to check symptoms is recommended to all personnel with suspected COVID-19 symptoms. Moreover, the city of Seoul has offered to officially support our project by placing a banner linked to our application on their website. Further discussions with government departments are underway.

Limitations and Conclusion

There are limitations to this study. The algorithm was based on KCDC guidelines, review of a few publications, and the expert opinions of 6 physicians. Evidence-based guidelines should be developed using data from patients with COVID-19. Insufficient data were available to develop an evidence-based model at the time, but additional studies developing statistical models using the newly acquired patient data will be followed. Additionally, one of the most important variables in the algorithm is body temperature. The typical household thermometer, especially the noncontact infrared thermometers, may be inaccurate, providing false guidance to users [28-30]. Another limitation of this study is that it does not provide any validation results on whether the algorithm was successful in patient triaging or increased efficiency in resource allocation. Concerns about the users' privacy led to the development of an application that does not collect or send any information provided by the user. Thus, no data were available for use in validating the efficacy of the model. Further, there is regional bias in the usage pattern since the application was developed and advertised only in Korea. However, many people in other countries have used the app since it supports multiple languages. Moreover, the usage pattern was not analyzed temporally because it was difficult to access all the temporal marks that can affect usage patterns.

An expert-opinion–based algorithm and application for patient screening and guidance can be beneficial in a circumstance where there is insufficient information on a novel disease and medical resources are limited.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Geographical distribution of app users.

XSL•FO

RenderX

[DOCX File , 150 KB - jmir_v22i11e19665_app1.docx]

Multimedia Appendix 2
List of COVID-19–related apps.
[DOCX File , 22 KB - jmir_v22i11e19665_app2.docx]

References

1. WHO Timeline - COVID-19. World Health Organization. 2020 Apr 27. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-05-01]
2. Coronavirus Disease-19 (COVID-19): Public Advice. KCDC. 2020. URL: [accessed 2020-05-01]
3. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. N Engl J Med 2020 Feb 20;382(8):692-694. [doi: 10.1056/NEJMp2000929] [Medline: 31978293]
4. Sandrock CE. Care of the critically ill and injured during pandemics and disasters: groundbreaking results from the Task Force on Mass Critical Care. Chest 2014 Oct;146(4):881-883. [doi: 10.1378/chest.14-1900] [Medline: 25144140]
5. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
6. Coronavirus Infectiousness-19 Response Guidelines for Local Governments Internet. Central Anti-Disaster Countermeasures Headquarters and Central Accident Control Headquarters. 2020. URL: [accessed 2020-06-03]
7. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
8. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore-Current Experience: Critical Global Issues That Require Attention and Action. JAMA 2020 Apr 07;323(13):1243-1244. [doi: 10.1001/jama.2020.2467] [Medline: 32077901]
9. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
10. Kim ES, Chin BS, Kang CK, Kim NJ, Kang YM, Choi J, Korea National Committee for Clinical Management of COVID-19. Clinical Course and Outcomes of Patients with Severe Acute Respiratory Syndrome Coronavirus 2 Infection: a Preliminary Report of the First 28 Patients from the Korean Cohort Study on COVID-19. J Korean Med Sci 2020 Apr 06;35(13):e142 [FREE Full text] [doi: 10.3346/jkms.2020.35.e142] [Medline: 32242348]
11. Flutter - Beautiful native apps in record time. Google, LLC. URL: https://flutter.dev [accessed 2020-05-11]
12. COVID-19 CheckUp. URL: https://ncov.docl.org [accessed 2020-10-21]
13. Google Analytics for Small Businesses. Google, LLC. URL: https://analytics.google.com/analytics/web/provision/#/provision [accessed 2020-05-01]
14. Google Analytics. Google, LLC. URL: https://developers.google.com/analytics/devguides/reporting/core/v4 [accessed 2020-05-01]
15. Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE Internet. Johns Hopkins University Center for Systems Science and Engineering. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-06-08]
16. National Information Society Agency. 2019 The Report on the Digital Divide. South Korea: Ministry of Science and ICT; 2020:22.
17. Lee D, Lee J. Testing on the move: South Korea's rapid response to the COVID-19 pandemic. Transportation Research Interdisciplinary Perspectives 2020 May;5:100111. [doi: 10.1016/J.TRIP.2020.100111]
18. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ 2020 Mar 23;368:m1165. [doi: 10.1136/bmj.m1165] [Medline: 32205334]
19. Wang Y, Liu Y, Liu L, Wang X, Luo N, Li L. Clinical Outcomes in 55 Patients With Severe Acute Respiratory Syndrome Coronavirus 2 Who Were Asymptomatic at Hospital Admission in Shenzhen, China. J Infect Dis 2020 May 11;221(11):1770-1774 [FREE Full text] [doi: 10.1093/infdis/jiaa119] [Medline: 32179910]
20. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci China Life Sci 2020 May;63(5):706-711 [FREE Full text] [doi: 10.1007/s11427-020-1661-4] [Medline: 32146694]
21. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiology of COVID-19 Among Children in China. Pediatrics 2020 Jun 16;145(6):e. [doi: 10.1542/peds.2020-0702] [Medline: 32179660]
22. DOCL Outcome Prediction Platform (Doctors on the Cloud). World Health Organization. URL: https://digitalatlas.who.int/en/covid-19/projects/1274/published [accessed 2020-06-08]
23. Cases in Korea. KCDC. URL: http://ncov.mohw.go.kr/en/bdBoardList.do?brdId=16&brdGubun=161&dataGubun=&ncvContSeq=&contSeq=&board_id=&gubun=
24. COVID-19 Screening Tool. Apple Inc. URL: https://covid19.apple.com/screening/ [accessed 2020-10-22]

25.   COVID Symptom Study - Help slow the spread of COVID-19. ZOE. URL: https://covid.joinzoe.com [accessed 2020-06-08]

26.   Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

27.   Self-Check Mobile App Instructions. The Ministry of Health and Welfare. URL: http://ncov.mohw.go.kr/selfcheck/ [accessed 2020-10-22]

28.   Hayward G, Verbakel J, Ismail F, Edwards G, Wang K, Fleming S, et al. Non-contact infrared versus axillary and tympanic thermometers in children attending primary care: a mixed-methods study of accuracy and acceptability. Br J Gen Pract 2020 Mar 23;70(693):e236-e244. [doi: 10.3399/bjgp20x708845]

29.   Niven DJ, Gaudet JE, Laupland KB, Mrklas KJ, Roberts DJ, Stelfox HT. Accuracy of Peripheral Thermometers for Estimating Temperature. Ann Intern Med 2015 Nov 17;163(10):768. [doi: 10.7326/m15-1150]

30.   Chung W, Chen C. Evaluation of performance and uncertainty of infrared tympanic thermometers. Sensors (Basel) 2010 Mar 31;10(4):3073-3089 [FREE Full text] [doi: 10.3390/s100403073] [Medline: 22319287]

Abbreviations

EMR: electronic medical record
ICT: information and communications technology
KCDC: Korean Center for Disease Control and Prevention
NSAID: nonsteroidal anti-inflammatory drug

XSL•FO
RenderX

Original Paper

The Hidden Pandemic of Family Violence During COVID-19: Unsupervised Learning of Tweets

Jia Xue[1,2], PhD; Junxiang Chen[3], PhD; Chen Chen[4], PhD; Ran Hu[1], MA, MSW; Tingshao Zhu[5], PhD

[1]Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

[2]Faculty of Information, University of Toronto, Toronto, ON, Canada

[3]School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[4]Middleware System Research Group, University of Toronto, Toronto, ON, Canada

[5]Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Corresponding Author:
Jia Xue, PhD
Factor-Inwentash Faculty of Social Work
University of Toronto
246 Bloor St W
Toronto, ON, M5S 1V4
Canada
Phone: 1 416 946 5429
Email: jia.xue@utoronto.ca

Abstract

Background: Family violence (including intimate partner violence/domestic violence, child abuse, and elder abuse) is a hidden pandemic happening alongside COVID-19. The rates of family violence are rising fast, and women and children are disproportionately affected and vulnerable during this time.

Objective: This study aims to provide a large-scale analysis of public discourse on family violence and the COVID-19 pandemic on Twitter.

Methods: We analyzed over 1 million tweets related to family violence and COVID-19 from April 12 to July 16, 2020. We used the machine learning approach Latent Dirichlet Allocation and identified salient themes, topics, and representative tweets.

Results: We extracted 9 themes from 1,015,874 tweets on family violence and the COVID-19 pandemic: (1) increased vulnerability: COVID-19 and family violence (eg, rising rates, increases in hotline calls, homicide); (2) types of family violence (eg, child abuse, domestic violence, sexual abuse); (3) forms of family violence (eg, physical aggression, coercive control); (4) risk factors linked to family violence (eg, alcohol abuse, financial constraints, guns, quarantine); (5) victims of family violence (eg, the LGBTQ [lesbian, gay, bisexual, transgender, and queer or questioning] community, women, women of color, children); (6) social services for family violence (eg, hotlines, social workers, confidential services, shelters, funding); (7) law enforcement response (eg, 911 calls, police arrest, protective orders, abuse reports); (8) social movements and awareness (eg, support victims, raise awareness); and (9) domestic violence–related news (eg, Tara Reade, Melissa DeRosa).

Conclusions: This study overcomes limitations in the existing scholarship where data on the consequences of COVID-19 on family violence are lacking. We contribute to understanding family violence during the pandemic by providing surveillance via tweets. This is essential for identifying potentially useful policy programs that can offer targeted support for victims and survivors as we prepare for future outbreaks.

(J Med Internet Res 2020;22(11):e24361) doi:10.2196/24361

Introduction

As seen in the case of Ebola, epidemics increase the rates of domestic violence [1]. The World Health Organization declared COVID-19 a pandemic on March 11, 2020. To effectively control the spread of the disease, many countries have adopted rigorous measures to limit mobility, such as social distancing, stay-at-home orders (sheltering in place), closure of nonessential

business, travel restrictions, and quarantine. Even though these measures are useful for infection control [2], they bring a series of negative social consequences, such as psychological stress [3-5], unemployment [6], ageism [7], and increased rates of violence against women and children [8-11]. Since these rigorous measures overlap with many of the intervention strategies for family violence [2], they are likely to increase the vulnerability of victims of family violence (including intimate partner violence [IPV]/domestic violence, elder abuse, and child abuse), by increasing exposure to an exploitative relationship, reducing options for support [10], economic stress [12], and alcohol abuse [13,14]. For example, isolation limits social contact with families and social services, and thus may facilitate family violence and prevent victims from seeking help [15-17]. During the COVID-19 quarantine, the home becomes a dangerous place for victims while individuals are living in forced close quarters [18]. In addition, mental health exacerbated by social isolation increases the likelihood of locking victims of domestic violence in an unsafe home environment and increases their vulnerability [19]. UNICEF [20] reports that school closures increased child (sexual) abuse and neglect during the Ebola epidemic. It is also important to note that child abuse and domestic violence are likely to co-occur when isolated at home [21,22]. During the COVID-19 pandemic, scholars have suggested that new forms of family violence may occur; for example, abusers may threaten to infect their family members with the virus [23].

In many countries, the reported cases of and service needs related to family violence dramatically increased since quarantine measures came into effect [18]. For example, calls to domestic violence hotlines have risen by 25%, and the number of Google searches for family violence–related help during the outbreak has been substantial [24]. According to National Domestic Violence Hotline representatives in the United States, abusers are attempting to isolate victims from resources and unleashing more violence by enforcing COVID-19 social distancing measures [25]. In the United Kingdom, calls to the Domestic Violence Helpline increase by 25% in the first week after the lockdown measures were implemented [26]. In China, domestic violence increased three times in Hubei Province during the lockdown [27]. There was a 10.2% increase in domestic violence calls in the United States during the COVID-19 pandemic [28]. These reports illustrate that existing COVID-19 intervention measures (eg, living in a closed space with abusers for a long period) may profoundly impact victims and survivors of family violence. According to Bradbury-Jones and Isham [8], "domestic violence rates are rising, and they are rising fast" (p 2047). Data on family violence during the pandemic are still scarce [29], and there is a need for further research.

We cannot capture the impact of COVID-19 on family violence without adequate surveillance [30]. Enhanced surveillance provides an understanding of the impact and risk factors associated with COVID-19, which is essential for developing policy programs to respond and mitigate adverse effects and offer targeted support for victims and survivors [30]. Eysenbach [31] defined infodemiology and infoveillance as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy. Infodemiology data can be collected and analyzed in near real time" (p 1). According to Eysenbach's framework, the automated analysis of unstructured data related to family violence and COVID-19 is an application of an infoveillance study. Understanding public discussions can assist governments and public health authorities in navigating the outbreak [32].

During the implementation of social isolation measures, social media should be leveraged to raise public awareness and share best practices (eg, bystander approaches, supportive statements, obtaining help on behalf of a survivor) [2], and provide support [33]. Twitter is a real-time network that allows users from across the globe to communicate via public and private messages, organized chronologically on a given user's account. Existing studies have confirmed Twitter's role in connecting practitioners and clients [34-36]. Researchers have used Twitter data to examine the nature of domestic violence [37-39]. A significant number of studies describe Twitter hashtag #MeToo as a phenomenal tool for disclosing experiences of sexual harassment, and more importantly, to ignite a widespread social campaign or political protest on social media. Modrek and Chakalov [40] examined tweets containing #MeToo in the United States and supported the role of machine learning methods in understanding the widespread sexual assault self-revelations on Twitter. Recently, Twitter has become a valuable source for understanding user-generated COVID-19 content and activities in real time [41,42].

Aim of the Study

There is a lack of data on the COVID-19 pandemic as it relates to family violence [43]. This study aims to provide a large-scale analysis of public discourse on family violence and COVID-19 on Twitter using machine learning techniques to fill this gap. The research questions are as follows: (1) what contents are discussed relating to family violence and COVID-19? and (2) what themes are identified relating to family violence and COVID-19? The study offers a new perspective on the impact, risks factors, and continuing social support services during the pandemic for family violence.

Methods

This study employed an observational design and followed the pipeline developed by the authors [44], including sampling, data collection, preprocessing of raw data, and data analysis.

Sampling and Data Collection

Our COVID-19 data set used a list of COVID-19–relevant hashtags as search terms to randomly collect tweets from Twitter between April 12 to July 16, 2020 [44] (Multimedia Appendix 1). Twitter Developer's Python code was used to access the Twitter API to collect tweets. As shown in Figure 1, our data set included a total of 274,501,992 tweets during the study period, of which 186,678,079 were in English. We sampled tweets using keywords such as "domestic violence," "intimate partner violence," "family violence," "violence against women," "gender-based violence," "child abuse," "child maltreatment,"

"elder abuse," and "IPV." The final data set comprised 1,015,874 tweets.

Figure 1. Tweets preprocessing chart.
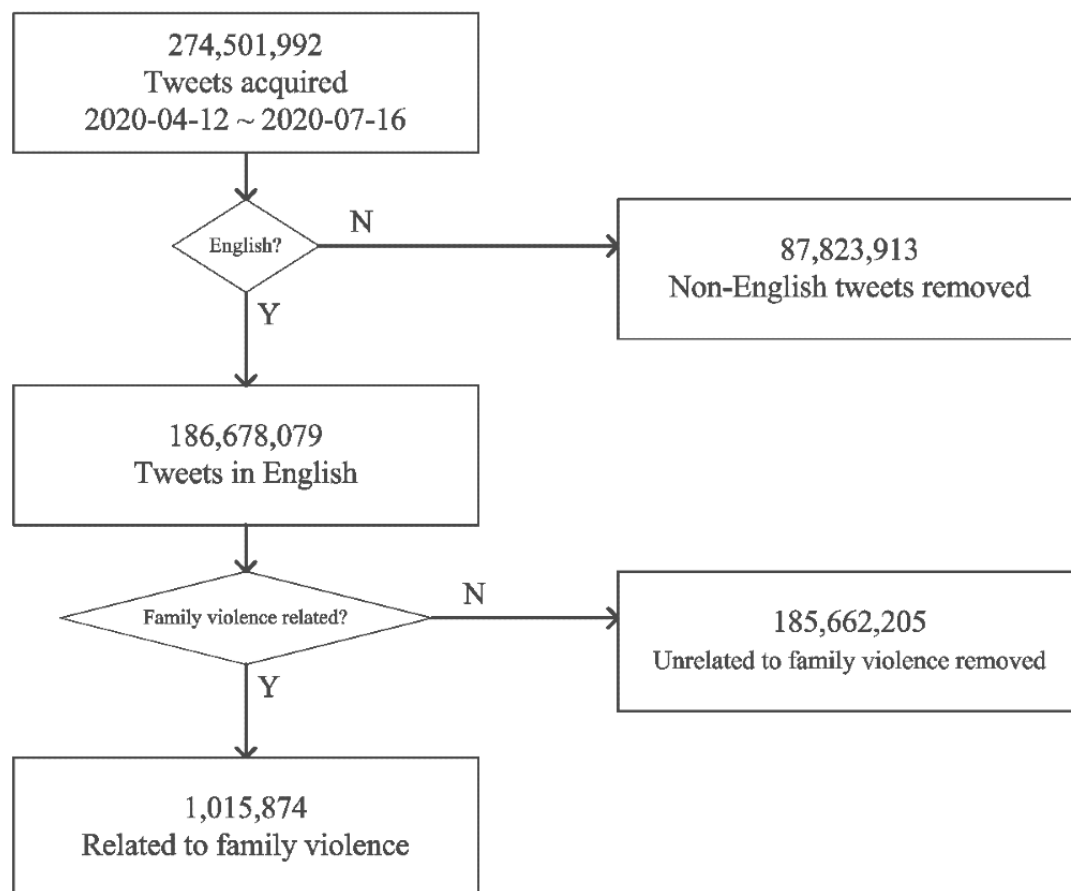
Preprocessing the Raw Data

We used Python to clean the data and remove the following items because they did not contribute to the semantic meaning of the tweets: the hashtag symbol, URLs, @users, special characters, punctuations, and stop-words [38,39,44,45].

Unsupervised Machine Learning

We used a machine learning approach, Latent Dirichlet Allocation (LDA) [46], to analyze a corpus of unstructured text. LDA was a generative statistical model that regards a corpus of text (tweets) as a mixture of a small number of latent topics. Each latent topic was assigned with a set of linguistic units (eg, single words or a pair of words) counted by the algorithm. These linguistic units with high frequency were likely to co-occur and form into different latent topics. With the LDA model, the distribution of topics in documents can be inferred. LDA assumes a generative process describing how the documents are created, such that we can infer or reverse engineer the topic distributions. The generative process of LDA for M documents, each of which has a length of N_i, is given as:

1. Choose $\theta_i \sim$ Dir (α), with $i \in \{1,\ldots,M\}$.
2. Choose $\varphi_k \sim$ Dir (β), with $k \in \{1,\ldots,K\}$.
3. For the j-th linguistic unit in the i-th document with $i \in \{1,\ldots,M,$ and $j \in \{1,\ldots,N N_i\}$
   a. Choose $z_{i,j} \sim$ Multinomial (θ_i)
   b. Choose $w_{i,j} \sim$ Multinomial ($\varphi_{z_{i,j}}$)

Multimedia Appendix 2 presents the definitions of these notations. With the generative process described above, the distributions of the topics can be inferred using the Python package *genism*.

Results

We analyzed 1,015,874 tweets mentioning family violence and COVID-19 in Twitter posts. We identified 50 latent topics and frequently mentioned pairs of words (bigrams) for each topic. We further categorized these 50 identified common topics into 9 themes and 33 topics (Table 1). Table 1 presents commonly co-occurring bigrams and examples of representative tweets under each identified theme and topic.

Table 1. Themes, topics, commonly co-occurring terms, and examples of tweets about domestic violence and COVID-19.

Themes and topics	Terms	Tweet example
Increased vulnerability: COVID-19 and family violence		
Rising rates	violence increase, violence spike, violence risen, abuse up	"…seeking shelter at home, rates of domestic violence and abuse have skyrocketed. Further, women and girls at a high risk for trafficking…"
DV[a] reports	violence reports, reports surge	"Several countries saw spikes in domestic violence reports."
Hotline calls increased	crisis line, violence hotline, abuse hotline, calls increased, calls help	"…works with domestic violence survivors is seeing a spike in emergency shelter capacity and crisis line calls during the coronavirus pandemic."
Homicide	murder domestic, violence homicide, murder wife	"Lots of social behaviours and issues are sadly coming to the fore during this pandemic…domestic violence where in the UK 2 women are killed…"
Suicide	suicide increased, suicide domestic, abuse suicide	"NY Gov. Cuomo: Suicides and increased domestic violence worth prices of coronavirus lockdown."
(Mental) health	mental health, mental abuse, mental illness, abuse depression	"Longer shutdowns = even more domestic violence, more substance abuse, more lonliness. More mental health symptoms."
Types of family violence during COVID-19		
Child abuse/maltreatment	assault child, children suffering, rape child, FGM[b] child	"If we're going to talk about quarantine, please don't forget that children are also in danger."
Domestic violence	abusive partners, risk family, family impacted	"With #StayHome orders, many women are left in isolation with abusive partners, unable to access life-saving resources and support systems."
Sexual violence	sexual assault, abuse rape, marital rape, rape incest, sexual abuse	"Thousands across the country are infected … shelters for rough sleepers, sexual abused have no place to go!"
Forms of family violence during COVID-19		
Physical aggression	stop hitting, physical domestic, physical abuse, physical violence	"…violence against women and girls has risen dramatically. My fear is more women and girls will die from physical violence than #Covid19."
Coercive control	power control, forced stay, coercive control, run away	"Domestic violence is about power and control. Abusers use more coercive control tactics surrounding the #covid19 pandemic to continue to maintain power and control over their partner."
Risk factors linked to family violence during COVID-19		
Drug abuse	overdose domestic, drug abuse, violence drug, addiction domestic	"@CaesarPodcast As I've noted for years, the history of domestic violence and drug abuse was enough to make him a prime suspect."
Alcohol abuse	liquor shops, violence alcoholism, suicide alcohol, violence alcohol	"March 2020 saw a surge in 'reported' cases of domestic violence. Alcoholism increases chances of abuse manifold on women's and children."
Financial constrain	violence unemployment, financial ruin, job lost, violence financial	"Take measures to stop domestic violence, which is on the rise due to lock down pressures. No work and no income. Women are facing double burden of providing food for the family and also facing the violence created by the frustrations."
Guns	gun control, violence guns, gun laws, gun violence, violence gun	"During this COVID-19 pandemic we are seeing a rise in gun sales, and a drastic increase in domestic violence cases nationwide."
Trafficking	human trafficking, sex trafficking, trafficking domestic	"This is part of our anti-trafficking #COVID19 response and our new DV response and led by CM @AbbieKamin, @hawctalk."
COVID-19 related	people stuck, violence covid19, unsafe home, abuse quarantine	"During the lockdown domestic violence happens, because the stress and also by the fact that coworkers/friends won't see the bruises."
Victims of family violence during COVID-19		
LGBTQ[c]	trans people, trans women, men men, lesbian couples	"COVID-19 has serious consequences for cis and trans women everywhere including higher risks a result of …the rise in domestic violence."
Women and women of color	women disproportionately affected, beat wife, black women, female victims, women die	"COVID-19 induced isolation and quarantine disproportionately affect women and girls. Around the world, there has been an increase in sexual and gender-based violence during COVID-19."
Refugee women	refuge domestic, charity refuge, violence refuges	"Refugee women are at greater risk for gender-based violence during the COVID-19 lockdown #WorldRefugeeDay https://t.co/iUPeae1vo8"
Children	violence child, child abusers, abuse child	"Evidence shows that violence against children is increasing due to #COVID19 lockdown. The pandemic shouldn't create another pandemic of torture and Rights abuse against children."

XSL•FO
RenderX

Themes and topics	Terms	Tweet example
Social services for victims of family violence during COVID-19		
Hotline numbers called	877 863, 863, 6338, 1 800, 800 799, 799 7233, 799 safe, hotline 800, 800 273, 273 8255	"Domestic violence help is available during #COVID19. Call the @ndvh hotline at 1-800-799-7233, text LOVEIS to 22522, or log on to chat at https://t.co/KhIhLk0fq3. You are not alone. https://t.co/qKRP8i3wbr."
Resources	social workers, social service, safety plan, crisis center, limited access, service open	"Risk of abuse through coercive control's, addictions'; mental health. No social work input is disappointing https://t.co/SbRhrpRlIE."
Shelters	women shelter, violence shelters, seeking shelters, shelters open	"Shelters have formed a GBV safety plan as victims of domestic violence are likely to be forced to stay at home with an abuser for longer periods due to the lockdown. https://t.co/wUQPsKeC4L."
Funding	funding support, consider donating, raise funds	"…The beauty industry supports victims of dv during the COVID-19 by donating products to women's shelters and DV service organizations."
Social media	visit website, violence website, retweet help	"…National Domestic Violence Helpline is there for victims, but it's not for men. That is apparent from their website. https://t.co/bhYUbfOZmv"
First responders: social workers	social workers, responding domestic, violence call	"Social workers would just get shot along with the wife or family, so who should society send to domestic violence calls? Cops say domestic violence calls are the most dangerous"
Law enforcement responses		
Law enforcement	contact police, protection orders, 911 calls, police arrest, legal aid, protective orders, local police	"Calls to local police departments are up in the last month. Help is available. Visit: https://t.co/naSNuaDeP3 for a list of local resources. https://t.co/C1KeCOxKGh."
Reports of DV cases	cases reported, abuse reports, report abuse, increase reports	"Due to the #COVID19 lockdown, there are increasing reports that girls and young women are facing gender-based violence…https://bit.ly/3cgkfDt."
Social movements and awareness		
Support victims	help victims, support victims, campaign combat, zero tolerance, care victims, ask help, reach out, situation help, protect vulnerable	"COVID-19 Lockdown witnesses a global rise in Domestic Violence. Trapped at home with abusers at all hours, lacking privacy to reach out for help, and the sudden disappearance of regular support systems, has isolated individuals, specifically women and children, in violence. https://t.co/80Az37rMvy."
Awareness	awareness domestic, raise awareness, help raise, awareness month, raising awareness, spread awareness, assault awareness	"Women share horrific photos of injuries to raise awareness of domestic violence as 14 are killed since start of lockdown. pics of their horrendous injuries to raise awareness of dv, as killings in the home DOUBLED during the first three weeks of lockdown in the UK. https://t.co/8nOczKoWiu."
Domestic violence–related news		
Personnel and events	Johnny Depp, Tracy McCarter, Rikers Island, Melissa DeRosa, Keith Ellison, Chris Brown, Antonio Guterres, Alexandra McCabe, Robert Goforth, Tara Reade, Breann Leath, California's bail rules	"Today, Secretary to the Governor Melissa DeRosa issued a report to Governor Cuomo outlining the COVID-19 Domestic Violence Task Force's initial recommendations to reimagine New York's approach to services for domestic violence survivors. https://t.co/VpMJEd7Njc."

[a]DV: domestic violence.

[b]FGM: female genital mutilation.

[c]LGBTQ: lesbian, gay, bisexual, transgender, and queer or questioning.

Increased Vulnerability: COVID-19 and Family Violence

Tweets mentioning rising rates of domestic violence as a consequence of COVID-19 were frequent, with popular bigrams like "violence increased," "violence higher," "rising violence," and "violence skyrocketing." Increases in hotline calls and reports of family violence were also influenced by the ongoing COVID-19 pandemic (eg, calls increased, calls help, reports surge). A representative tweet indicated, "a Miami Valley nonprofit agent is seeing a spike in crisis line calls during the pandemic." Other consequences of the pandemic include homicides related to domestic violence and mental health issues (eg, depression, mental abuse).

Types of Family Violence During COVID-19

Findings showed that several types of family violence were mentioned together in a single tweet alongside terms related to COVID-19, such as "child abuse/maltreatment" (eg, assault child, rape child), "domestic violence" (eg, abusive partners,

violence partners), and "sexual violence" (eg, sexually assault, marital rape).

Forms of Family Violence During COVID-19

Two primary forms of family violence were discussed on Twitter during the COVID-19: "physical aggression" (eg, physically hurt, stop hitting) and "coercive control" (eg, power control, forced stay). The latter is demonstrated by this example: "…abusers may use more coercive control tactics surrounding the #covid19 pandemic to continue to maintain power and control over their partner."

Risk Factors Linked to Family Violence During COVID-19

We found that the rising rate of domestic violence was associated with risk factors: "drug abuse," "alcohol abuse," "financial constraints" (eg, job loss, loss income), "guns," "trafficking," and "COVID-19 related" (eg, lockdown, stuck home, quarantine). Sample tweets include "March 2020 saw a surge in reported cases of domestic violence. Alcoholism increases chances of abuse manifold on women and children…" and "During the lockdown, domestic violence happens because the coworkers/friends can't see the bruises."

Victims of Family Violence During COVID-19

Tweets designated the LGBTQ (lesbian, gay, bisexual, transgender, and queer or questioning) community, women, women of color, refugee women, and children as victims of family violence during the COVID-19. Popular words in describing the victims and survivors of family violence included "trans people," "lesbian couples," "women disproportionately affected," "beat wife," "black women," "female victims," "refuge domestic," "charity refuge," "violence child," "child abusers," and "abuse child."

Social Services for Victims of Family Violence During COVID-19

Social services for victims of family violence was a prominent theme discussed by Twitter users during the pandemic, as indicated by the high frequency mentions of hotline numbers. Resources, shelters, funding support, and visiting websites on family violence were also frequently mentioned in tweets. In addition, confidential services, safety plans, and limited access were representative topics identified in the sampled tweets. Social workers' safety was tweeted as a salient topic in our data set: "…domestic violence cases are just asking for a lot of social workers to get shot and killed" and "Has anyone actually asked social workers how willing they are to go on domestic violence calls…?"

Law Enforcement Responses

With the rising rates of family violence during the pandemic, reports of domestic violence cases (eg, cases reported, abuse reports, violence reports, increase reports, and reported increase) were a salient topic in the tweets. Police departments (eg, police officers, local police, police chief, 911 calls, contact police, police arrest) were the first responders on the front lines during increased domestic violence reports during COVID-19.

Social Movements and Awareness

Findings also identified social justice movements and awareness to support victims and survivors of family violence. Tweet content highlighted the advocacy of zero tolerance for domestic violence, indicated by popular bigrams such as "help victims," "campaign combat," "violence advocacy," "care victims," "raise awareness," and "awareness campaign" and sample tweets like "Women share horrific photos of injuries to raise awareness of domestic violence."

Domestic Violence–Related News

News events related to domestic violence cases during the pandemic were also identified, such as (1) American actor Johnny Depp's denial of domestic abuse allegations by ex-wife Amber Heard; (2) Tracy McCarter's murder charge for the fatal stabbing of her husband in Manhattan; (3) singer Chris Brown's arrest in Paris on allegations of rape; (4) Tara Reade's sexual assault allegations against Joe Biden; (5) Kentucky legislator Robert Goforth's arrest for 4th-degree domestic violence; and (6) death of police officer Breann Leath, who was shot on duty while responding to a domestic disturbance call.

News of solutions to help survivors of domestic violence were also frequently discussed in the sampled tweets. For example, the governor of the New York State Council on Women and Girls, Melissa DeRosa, created a task force to find innovative solutions to the domestic violence spike during the COVID-19 pandemic. United Nations chief Antonio Guterres called for measures to address the surge in domestic violence linked to lockdowns that were imposed by governments in responding to the COVID-19 pandemic.

The news article *Child abusers eligible for immediate release under California's new $0 cash bail emergency mandate* [47] has become a prominent topic due to the high volume of retweets. Given the new state rules, individuals arrested for child abuse will be released on $0 bail in California. The original tweet was posted by Bill Melugin (@billFOXLA) and had been retweeted almost 1000 times ("RT @BillFOXLA: Under California's new $0 cash bail rules, child abusers are now eligible for immediate release. San Bernardino County Sheriff @sheriffmcmahon tells me he had to release a felony child abuse suspect /w priors for domestic violence & child abuse immediately after arrest. @FOXLA").

Discussion

Principal Results

Our study employed a large-scale analysis of tweets on public discourse related to family violence on Twitter during the COVID-19 pandemic. The study's Twitter data consisted of a random selection of more than 1 million tweets mentioning family violence and COVID-19 from April 12 to July 16, 2020. The machine learning technique LDA was used to extract a high volume of co-occurring word pairs and topics related to family violence from unstructured tweets. The study contributes to the understanding of public discourse and concerns of family violence during the COVID-19 pandemic. We identified 9 themes from the analysis: (1) increased vulnerability: COVID-19 and family violence (eg, increasing rates, victims affected); (2)

types of family violence; (3) forms of family violence; (4) victims of family violence; (5) risk factors linked to family violence; (6) social services for victims of family violence; (7) law enforcement responses; (8) social movements and awareness; and (9) domestic violence–related news. The study adds to existing scholarship, where there is a lack of data on the COVID-19–domestic violence connection, or only anecdotal reports. Our findings contribute to understanding family violence during the pandemic by providing surveillance via tweets, which is essential to identify potentially effective policy programs in offering targeted support for victims and survivors and preparing for future outbreaks.

Twitter users have discussed who is at higher risk of family violence during the lockdown. Findings reveal a broader range of affected victims, such as the LGBTQ community. Salient tweets suggest that women and children are disproportionately affected by family violence that is consistent with the majority of the research in the field [38,48-51]. Violence against children has been associated with previous epidemics [6]. In addition, the sampled tweets suggest that domestic violence–related discussions focus on the support and protection of victims instead of interventions against abusers, consistent with one recent study using Twitter data for domestic violence research [39]. We find tweets mentioning family violence and COVID-19 have a limitation in primarily posting stories about male-to-female violence [37] even though other patterns of violence exist, including female-to-male, male-to-male, and bidirectional IPV [52].

Tweets about family violence and COVID-19 during the lockdown mentioned a range of risk factors associated with family violence during pandemics, such as drug abuse, alcohol abuse, financial constraints, guns, and trafficking. Our study reveals similar results with one recent report by Peterman and colleagues [10], who summarized that 9 main pathways that connect the COVID-19 pandemic and violence against women and children (ie, economic insecurity and poverty-related stress; quarantines and social isolation; disaster- and conflict-related unrest and instability; and inability to temporarily take shelter from abusive partners). For example, public discussions indicate that alcohol abuse continues to be a risk factor for family violence during stressful events [53]. Financial constraints (eg, financial ruin, lost jobs, economic collapse) due to COVID-19 create barriers for victims of family violence for help seeking [2]. Beland and colleagues [54] analyzed the Canadian Perspective Survey Series and found that financial worries due to COVID-19 contributed to increased family violence and stress. An increasing rate of domestic homicides identified in tweets suggests that guns are still a concern at home where family violence occurs. Specific COVID-19–related risk factors (eg, quarantines, social isolation) limit contact between victims of family violence and the outside world, trapping them at home with their abusers; these factors were indicated by the frequent use of words like "people stuck," "unsafe home," "people locked," and "abuse quarantine" on Twitter.

Multiagency integration of law enforcement responses (eg, protection orders, arrest), social services (eg, hotlines, shelters), and social movements and awareness are recommended to address domestic violence and support victims [55]. Social

services (including deployment of social work practitioners, therapists, etc) for cases of domestic violence must be resourced during the pandemic. Due to the mobility restriction, a lack of informal support, such as that from family, friends, coworkers, further contributes to increased rates of family violence during the pandemic. Thus, it is more crucial than ever for victims to access voluntary sector practitioners' support during the COVID-19 pandemic [8]. Our results provide evidence that some agencies continued to deliver services during the pandemic. For example, several hotline numbers in the United States have been frequently mentioned during the pandemic, such as "Illinois Domestic Violence Hotline, 877-863-6338 (877-TO END DV)," "National Suicide Prevention Lifeline, 800-273-8255 (US)," "National Domestic Violence Hotline, 800-799-SAFE (7233) (US)," "National Sexual Assault Telephone Hotline, 800-656-HOPE (4673) (US)," and "Loveisrespect, Text LOVEIS to 22522 (US)." We also identified popular hotline numbers from the United Kingdom, such as Mind the Mental Health Charity (Mind Infoline: 0300-123-3393), the National Stalking Helpline (0808-802-0300), and the National Domestic Abuse Helpline (0808-2000-247). However, a commentary in the *Canadian Medical Association Journal* raises concerns about family violence support using videoconference or telemedicine settings where the abusers can be present [56]. Abusers can coercively control victims-survivors' use of mobile phones to access hotline support. Therefore, further evidence is needed to indicate whether the services fulfill their roles.

Twitter conversations about highly publicized domestic violence cases were significant. News about Hollywood star Johnny Depp's denial of abuse allegations when he was accused of domestic violence against his ex-wife Amber Heard was a prominent topic in the sampled tweets. Our results show public discussions of high-profile cases of domestic violence (eg, athletes arrested for domestic violence), consistent with previous studies. Cravens et al [37] used qualitative content analysis to examine the factors that influence IPV victims to leave an abusive relationship using 676 tweets related to #whyIstayed and #whyIleft. Xue et al [38] analyzed 322,863 tweets about domestic violence and found that high-profile cases such as Greg Hardy's domestic violence case are prominent. These studies consistently show that Twitter continues to be a source of news coverage on current events for domestic violence, even during the COVID-19 pandemic.

Limitations

There are a number of limitations to this study that must be acknowledged. First, Twitter data reveal insights from Twitter users and thus does not represent the entire population's opinions. Despite this shortcoming, our study provides one of the first large-scale analysis of tweets using real-time data to identify the impact of COVID-19 on family violence. Second, we did not include non-English tweets in the analysis. Future studies should carry out analyses on non-English tweets regarding the impact of COVID-19 on family violence. Third, even though our collected data cover 90 days of the outbreak since April 12, 2020, discussion patterns may evolve as the COVID-19 situation continues to change over time. Fourth, the search terms used in the study mostly reflect terminology used

by professionals rather than victims when discussing family violence. For example, one study examined how child abuse victims post their stories on social media and found that the victims rarely use explicit words to describe their experiences [57]. Thus, this study may be limited in capturing victims' opinions. To protect Twitter users' privacy and anonymity, we did not examine the sample's sociodemographic characteristics. It remains unknown whether the collected tweets were from victims, abusers, organizations, etc. It is also possible that abusers may prevent victims from reaching out for help on social media [9]. Future studies could consider sampling tweets from victims of family violence to further examine the impact of COVID-19.

Conclusion

As seen in our large-scale tweets data set, people have been actively discussing family violence in the context of COVID-19. We identified 9 themes and 33 topics relating to family violence and COVID-19. The findings demonstrate that Twitter can serve as a platform for real-time and large-scale surveillance of family violence by offering an understanding of the people who are discussing the impact and risk factors associated with COVID-19, which is essential for developing policy programs for supporting victims and survivors. This study provides insights for professionals who work with victims and survivors of family violence to develop a social network–based support system for informal and formal help when conventional in-person support services become unavailable during future outbreaks.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hashtags used as data collection search terms.
[DOCX File , 13 KB - jmir_v22i11e24361_app1.docx]

Multimedia Appendix 2
Notations for Latent Dirichlet Allocation (LDA).
[DOCX File , 13 KB - jmir_v22i11e24361_app2.docx]

References

1. Everything on her shoulders: rapid assessment on gender and violence against women and girls in the Ebola outbreak in Beni, DRC. International Rescue Committee. 2019 Mar 15. URL: https://www.rescue.org/report/ everything-her-shoulders-rapid-assessment-gender-and-violence-against-women-and-girls-ebola [accessed 2020-07-11]
2. van Gelder N, Peterman A, Potts A, O'Donnell M, Thompson K, Shah N, GenderCOVID-19 working group. COVID-19: Reducing the risk of infection might increase the risk of intimate partner violence. EClinicalMedicine 2020 Apr 11:100348 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100348] [Medline: 32292900]
3. Galea S, Merchant RM, Lurie N. The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
4. Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
5. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12) [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
6. Roje Đapić M, Buljan Flander G, Prijatelj K. Children Behind Closed Doors Due to COVID-19 Isolation: Abuse, Neglect and Domestic Violence. Arch Psychiatry Res 2020 May 15;56(2):181-192. [doi: 10.20471/dec.2020.56.02.06]
7. Xiang X, Lu X, Halavanau A, Xue J, Sun Y, Lai P, et al. Modern Senicide in the Face of a Pandemic: An Examination of Public Discourse and Sentiment about Older Adults and COVID-19 Using Machine Learning. J Gerontol B Psychol Sci Soc Sci 2020 Aug 12:A [FREE Full text] [doi: 10.1093/geronb/gbaa128] [Medline: 32785620]
8. Bradbury-Jones C, Isham L. The pandemic paradox: The consequences of COVID-19 on domestic violence. J Clin Nurs 2020 Jul 22;29(13-14):2047-2049 [FREE Full text] [doi: 10.1111/jocn.15296] [Medline: 32281158]
9. Campbell AM. An increasing risk of family violence during the Covid-19 pandemic: Strengthening community collaborations to save lives. Forensic Science International: Reports 2020 Dec;2:100089. [doi: 10.1016/j.fsir.2020.100089]
10. Peterman A, Potts A, O'Donnell M, Thompson K, Shah N, Oertelt-Prigione S, et al. Pandemics and violence against women and children. Center for Global Development. 2020 Apr 1. URL: https://www.cgdev.org/publication/ pandemics-and-violence-against-women-and-children [accessed 2020-07-11]

11.   Joint Leaders' statement - Violence against children: A hidden crisis of the COVID-19 pandemic. World Health Organization. 2020 Apr 8. URL: https://tinyurl.com/y4hjedod [accessed 2020-07-11]

12.   Labor force statistics from the current population survey. US Bureau of Labor Statistics. URL: https://www.bls.gov/help/one_screen/ln.htm [accessed 2020-04-18]

13.   Colbert S, Wilkinson C, Thornton L, Richmond R. COVID-19 and alcohol in Australia: Industry changes and public health impacts. Drug Alcohol Rev 2020 Jul;39(5):435-440 [FREE Full text] [doi: 10.1111/dar.13092] [Medline: 32483882]

14.   Vora M, Malathesh BC, Das S, Chatterjee SS. COVID-19 and domestic violence against women. Asian J Psychiatr 2020 Jun 15;53:102227 [FREE Full text] [doi: 10.1016/j.ajp.2020.102227] [Medline: 32574942]

15.   Gelles RJ. An exchange/social control theory. In: Finkelhor D, Gelles RJ, Hotaling GT, Strauss MA, editors. The Dark Side of Families: Current Family Violence Research. Durham, NH: SAGE Publications; 1983:151-165.

16.   Gelles RJ, Strauss M. Determinants of violence in the family: Toward a theoretical integration. In: Contemporary Theories about the Family. New York, NY: Free Press; 1979.

17.   Hagan E, Raghavan C, Doychak K. Functional Isolation: Understanding Isolation in Trafficking Survivors. Sex Abuse 2019 Nov 28. [doi: 10.1177/1079063219889059] [Medline: 31777323]

18.   Salisbury H. Helen Salisbury: What might we learn from the covid-19 pandemic? BMJ 2020 Mar 17;368:m1087. [doi: 10.1136/bmj.m1087] [Medline: 32184276]

19.   Khalifeh H, Dean K. Gender and violence against people with severe mental illness. Int Rev Psychiatry 2010 Nov 03;22(5):535-546. [doi: 10.3109/09540261.2010.506185] [Medline: 21047165]

20.   COVID-19: Children at heightened risk of abuse, neglect, exploitation and violence amidst intensifying containment measures. UNICEF. 2020 Mar 20. URL: http://www.unicef.org/press-releases/covid-19-children-heightened-risk-abuse-neglect-exploitation-and-violence-amidst [accessed 2020-07-21]

21.   Humphreys KL, Myint MT, Zeanah CH. Increased Risk for Family Violence During the COVID-19 Pandemic. Pediatrics 2020 Jul 21;146(1):e20200982. [doi: 10.1542/peds.2020-0982] [Medline: 32317306]

22.   Vu NL, Jouriles EN, McDonald R, Rosenfield D. Children's exposure to intimate partner violence: A meta-analysis of longitudinal associations with child adjustment problems. Clin Psychol Rev 2016 Jun;46:25-33. [doi: 10.1016/j.cpr.2016.04.003] [Medline: 27136293]

23.   Emezue C. Digital or Digitally Delivered Responses to Domestic and Intimate Partner Violence During COVID-19. JMIR Public Health Surveill 2020 Jul 30;6(3):e19831 [FREE Full text] [doi: 10.2196/19831] [Medline: 32678797]

24.   Kottasová I, Di Donato V. Women are using code words at pharmacies to escape domestic violence during lockdown. CNN. 2020 Apr 6. URL: https://www.cnn.com/2020/04/02/europe/domestic-violence-coronavirus-lockdown-intl/index.html [accessed 2020-07-21]

25.   Godin M. How Coronavirus is Affecting Victims of Domestic Violence. Time. 2020 Mar 18. URL: https://time.com/5803887/coronavirus-domestic-violence-victims/ [accessed 2020-07-23]

26.   Kelly J, Morgan T. Coronavirus: Domestic Abuse Calls up 25% Since Lockdown, Charity Says. BBC News. 2020 Apr 6. URL: https://www.bbc.co.uk/news/uk-52157620 [accessed 2020-07-23]

27.   Fraser E. Impact of COVID-19 pandemic on violence against women and girls. Social Development Direct. 2020 Mar 16. URL: https://www.sddirect.org.uk/media/1881/vawg-helpdesk-284-covid-19-and-vawg.pdf [accessed 2020-07-23]

28.   Leslie E, Wilson R. Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. J Public Econ 2020 Sep;189:104241 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104241] [Medline: 32834179]

29.   Addressing violence against children, women and older people during the Covid-19 pandemic: Key actions. World Health Organization. 2020 Jun 18. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Violence_actions-2020.1 [accessed 2020-06-17]

30.   Chandan JS, Taylor J, Bradbury-Jones C, Nirantharakumar K, Kane E, Bandyopadhyay S. COVID-19: a public health approach to manage domestic violence is needed. The Lancet Public Health 2020 Jun;5(6):e309. [doi: 10.1016/s2468-2667(20)30112-2]

31.   Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

32.   Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]

33.   Boserup B, McKenney M, Elkbuli A. Alarming trends in US domestic violence during the COVID-19 pandemic. Am J Emerg Med 2020 Apr 28 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.077] [Medline: 32402499]

34.   Farhatiningsih L, Salamah U. Utilization of Twitter as communication strategy for issues management by government public relations practitioners. Semantic Scholar. 2018. URL: https://www.semanticscholar.org/paper/Utilization-of-Twitter-as-Communication-Strategy-by-Farhatiningsih-Salamah/aae209b09211a34636447ad6324f683c05b7b145 [accessed 2020-10-30]

35.   Lovejoy K, Waters RD, Saxton GD. Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. Public Relations Review 2012 Jun;38(2):313-318. [doi: 10.1016/j.pubrev.2012.01.005]

36. Park H, Rodgers S, Stemmle J. Analyzing health organizations' use of Twitter for promoting health literacy. J Health Commun 2013 Apr;18(4):410-425. [doi: 10.1080/10810730.2012.727956] [Medline: 23294265]

37. Cravens JD, Whiting JB, Aamar RO. Why I Stayed/Left: An Analysis of Voices of Intimate Partner Violence on Social Media. Contemp Fam Ther 2015 Sep 25;37(4):372-385. [doi: 10.1007/s10591-015-9360-8]

38. Xue J, Chen J, Gelles R. Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter. Violence and Gender 2019 Jun;6(2):105-114. [doi: 10.1089/vio.2017.0066]

39. Xue J, Macropol K, Jia Y, Zhu T, Gelles RJ. Harnessing big data for social justice: An exploration of violence against women‐related conversations on Twitter. Human Behav and Emerg Tech 2019 Jul 26;1(3):269-279. [doi: 10.1002/hbe2.160]

40. Modrek S, Chakalov B. The #MeToo Movement in the United States: Text Analysis of Early Twitter Conversations. J Med Internet Res 2019 Sep 03;21(9):e13837 [FREE Full text] [doi: 10.2196/13837] [Medline: 31482849]

41. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

42. Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public Priorities and Concerns Regarding COVID-19 in an Online Discussion Forum: Longitudinal Topic Modeling. J Gen Intern Med 2020 Jul 12;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]

43. Anurudran A, Yared L, Comrie C, Harrison K, Burke T. Domestic violence amid COVID-19. Int J Gynaecol Obstet 2020 Aug 15;150(2):255-256. [doi: 10.1002/ijgo.13247] [Medline: 32472696]

44. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]

45. Stone B, Dennis S, Kwantes P. Comparing methods for single paragraph similarity analysis. Top Cogn Sci 2011 Jan;3(1):92-122 [FREE Full text] [doi: 10.1111/j.1756-8765.2010.01108.x] [Medline: 25164176]

46. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]

47. Child abusers eligible for immediate release under California's new $0 cash bail emergency mandate. News Break. 2020 Apr 21. URL: https://www.newsbreak.com/news/1550710194577/child-abusers-eligible-for-immediate-release-under-californias-new-0-cash-bail-emergency-mandate [accessed 2020-07-21]

48. Aizer A. Poverty, Violence and Health: The Impact of Domestic Violence During Pregnancy on Newborn Health. J Hum Resour 2011;46(3):518-538 [FREE Full text] [doi: 10.1353/jhr.2011.0024] [Medline: 24839303]

49. Bradley NL, DiPasquale AM, Dillabough K, Schneider PS. Health care practitioners' responsibility to address intimate partner violence related to the COVID-19 pandemic. CMAJ 2020 Jun 01;192(22):E609-E610. [doi: 10.1503/cmaj.200634] [Medline: 32357996]

50. Fantuzzo J, Boruch R, Beriama A, Atkins M, Marcus S. Domestic violence and children: prevalence and risk in five major U.S. cities. J Am Acad Child Adolesc Psychiatry 1997 Jan;36(1):116-122. [doi: 10.1097/00004583-199701000-00025] [Medline: 9000789]

51. Tjaden P, Thoennes N. Prevalence, Incidence and Consequences of Violence Against Women: Findings from the National Violence Against Women Survey. National Institute of Justice, Centers for Disease Control and Prevention. 1998 Nov. URL: https://www.ncjrs.gov/pdffiles/172837.pdf [accessed 2020-07-21]

52. Hu R, Xue J, Lin K, Sun IY, Wang X. Bidirectional Intimate Partner Violence Among Chinese Women: Patterns and Risk Factors. J Interpers Violence 2019 Dec 03:886260519888523. [doi: 10.1177/0886260519888523] [Medline: 31795808]

53. Catalá-Miñana A, Lila M, Oliver A, Vivo J, Galiana L, Gracia E. Contextual Factors Related to Alcohol Abuse Among Intimate Partner Violence Offenders. Subst Use Misuse 2017 Feb 23;52(3):294-302. [doi: 10.1080/10826084.2016.1225097] [Medline: 27759488]

54. Beland LP, Brodeur A, Haddad J, Mikola D. IZA DP No. 13332: COVID-19, Family Stress and Domestic Violence: Remote Work, Isolation and Bargaining Power. IZA Institute of Labor Economics. 2020 Jun. URL: https://www.iza.org/publications/dp/13332/covid-19-family-stress-and-domestic-violence-remote-work-isolation-and-bargaining-power [accessed 2020-07-21]

55. Gulati G, Kelly BD. Domestic violence against women and the COVID-19 pandemic: What is the role of psychiatry? Int J Law Psychiatry 2020 Jul;71:101594 [FREE Full text] [doi: 10.1016/j.ijlp.2020.101594] [Medline: 32768101]

56. Bradley NL, DiPasquale AM, Dillabough K, Schneider PS. Health care practitioners' responsibility to address intimate partner violence related to the COVID-19 pandemic. CMAJ 2020 Jun 01;192(22):E609-E610 [FREE Full text] [doi: 10.1503/cmaj.200634] [Medline: 32357996]

57. Schwab-Reese L, Kanuri N, Cash S. Child Maltreatment Disclosure to a Text Messaging-Based Crisis Service: Content Analysis. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e11306 [FREE Full text] [doi: 10.2196/11306] [Medline: 30907745]

Abbreviations

LDA: Latent Dirichlet Allocation
LGBTQ: lesbian, gay, bisexual, transgender, and queer or questioning

XSL•FO

RenderX

Original Paper

Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation

Akhil Vaid[1], MD; Sulaiman Somani[1*], BSc; Adam J Russak[1,2*], MD; Jessica K De Freitas[1,3], BSc; Fayzan F Chaudhry[1,3], BSc; Ishan Paranjpe[1], BSc; Kipp W Johnson[3], PhD; Samuel J Lee[1], BSc; Riccardo Miotto[1,3], PhD; Felix Richter[1,3], PhD; Shan Zhao[1,4], MD, PhD; Noam D Beckmann[3], PhD; Nidhi Naik[1], MSc; Arash Kia[5,6], MD; Prem Timsina[5,6], PhD; Anuradha Lala[5,7], MD; Manish Paranjpe[8], BSc; Eddye Golden[1], MPH; Matteo Danieletto[1], PhD; Manbir Singh[1], BBA; Dara Meyer[3], MSc; Paul F O'Reilly[3,9,10], PhD; Laura Huckins[3,9,10], PhD; Patricia Kovatch[11], BSc; Joseph Finkelstein[5], MD, PhD; Robert M. Freeman[5,6], MBA, RN; Edgar Argulian[12,13], MD, MPH; Andrew Kasarskis[3,5,14,15], PhD; Bethany Percha[2], PhD; Judith A Aberg[2,16], MD; Emilia Bagiella[6,7], PhD; Carol R Horowitz[2,5], MD, MPH; Barbara Murphy[2], MD; Eric J Nestler[17,18], MD, PhD; Eric E Schadt[3,14], PhD; Judy H Cho[19], MD; Carlos Cordon-Cardo[20], MD, PhD; Valentin Fuster[7,12,13], MD, PhD; Dennis S Charney[21], MD; David L Reich[4], MD; Erwin P Bottinger[1,22], MD; Matthew A Levin[3,4], MD; Jagat Narula[12,13], MD, PhD; Zahi A Fayad[23,24], PhD; Allan C Just[25], PhD; Alexander W Charney[3,9,10], MD, PhD; Girish N Nadkarni[1,2,19], MD, MPH; Benjamin S Glicksberg[1,3], PhD

[1]The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

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

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[7]The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[8]Harvard Medical School, Boston, MA, United States

[9]The Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[10]The Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[11]Mount Sinai Data Warehouse, Icahn School of Medicine at Mount Sinai, New York, NY, United States

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

[13]Department of Cardiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[14]Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[15]Mount Sinai Data Office, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[16]Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[17]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[18]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[19]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[20]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[21]Office of the Dean, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[22]Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

[23]BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[24]Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[25]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States

*these authors contributed equally

Corresponding Author:
Benjamin S Glicksberg, PhD
The Hasso Plattner Institute for Digital Health at Mount Sinai
Icahn School of Medicine at Mount Sinai
770 Lexington St.

New York, NY
United States
Phone: 1 212 731 7078
Email: benjamin.glicksberg@mssm.edu

Abstract

Background: COVID-19 has infected millions of people worldwide and is responsible for several hundred thousand fatalities. The COVID-19 pandemic has necessitated thoughtful resource allocation and early identification of high-risk patients. However, effective methods to meet these needs are lacking.

Objective: The aims of this study were to analyze the electronic health records (EHRs) of patients who tested positive for COVID-19 and were admitted to hospitals in the Mount Sinai Health System in New York City; to develop machine learning models for making predictions about the hospital course of the patients over clinically meaningful time horizons based on patient characteristics at admission; and to assess the performance of these models at multiple hospitals and time points.

Methods: We used Extreme Gradient Boosting (XGBoost) and baseline comparator models to predict in-hospital mortality and critical events at time windows of 3, 5, 7, and 10 days from admission. Our study population included harmonized EHR data from five hospitals in New York City for 4098 COVID-19–positive patients admitted from March 15 to May 22, 2020. The models were first trained on patients from a single hospital (n=1514) before or on May 1, externally validated on patients from four other hospitals (n=2201) before or on May 1, and prospectively validated on all patients after May 1 (n=383). Finally, we established model interpretability to identify and rank variables that drive model predictions.

Results: Upon cross-validation, the XGBoost classifier outperformed baseline models, with an area under the receiver operating characteristic curve (AUC-ROC) for mortality of 0.89 at 3 days, 0.85 at 5 and 7 days, and 0.84 at 10 days. XGBoost also performed well for critical event prediction, with an AUC-ROC of 0.80 at 3 days, 0.79 at 5 days, 0.80 at 7 days, and 0.81 at 10 days. In external validation, XGBoost achieved an AUC-ROC of 0.88 at 3 days, 0.86 at 5 days, 0.86 at 7 days, and 0.84 at 10 days for mortality prediction. Similarly, the unimputed XGBoost model achieved an AUC-ROC of 0.78 at 3 days, 0.79 at 5 days, 0.80 at 7 days, and 0.81 at 10 days. Trends in performance on prospective validation sets were similar. At 7 days, acute kidney injury on admission, elevated LDH, tachypnea, and hyperglycemia were the strongest drivers of critical event prediction, while higher age, anion gap, and C-reactive protein were the strongest drivers of mortality prediction.

Conclusions: We externally and prospectively trained and validated machine learning models for mortality and critical events for patients with COVID-19 at different time horizons. These models identified at-risk patients and uncovered underlying relationships that predicted outcomes.

Introduction

Despite substantial, organized efforts to prevent disease spread, over 23 million people have tested positive for SARS-CoV-2 worldwide, and the World Health Organization has reported more than 800,000 deaths from the virus to date [1-4]. As a result of this pandemic, hospitals are being filled beyond capacity and face extreme challenges with regard to personnel staffing, personal protective equipment availability, and intensive care unit (ICU) bed allocation. Additionally, patients with COVID-19 demonstrate varying symptomatology, making safe and successful patient triaging difficult. While some infected patients are asymptomatic, others suffer from severe acute respiratory distress syndrome, experience multiorgan failure, or die [5-7]. Identification of key patient characteristics that govern the course of disease across large patient cohorts is important, particularly given its potential to aid physicians and hospitals in predicting disease trajectory, allocating essential resources effectively, and improving patient outcomes.

Prognostication with machine learning is poised to accomplish this [8]; however, efforts have been limited by small sample sizes, lack of generalization to diverse populations, disparities in feature missingness, and potential for bias [9]. Many predictive models have met with success; however, these models only consider demographics, clinical symptoms, or laboratory values rather than considering all these factors conjointly [10-17]. More recent studies have accounted for fundamental aspects of machine learning but are limited in scope [13,18-22]. These studies lack either temporal benchmarks, interhospital or prospective validation, systematic evaluation of multiple models, consideration of covariate correlations, or assessment of the impact of the imputed data. With these needs in mind, we report the development of a boosted decision tree–based machine learning model trained on electronic health records from patients confirmed to have COVID-19 at a single center in the Mount Sinai Health System (MSHS) in New York City to predict critical events and mortality. To assess both interhospital and temporal generalizability, we first externally validated this algorithm to four other hospital centers. We then

prospectively validated it on a new set of patients from all five hospitals. Finally, we performed a saliency analysis using SHAP (SHapley Additive exPlanation) values to identify the most important features used by this model for outcome prediction.

Methods

Clinical Data Sources

Patient data were obtained from five hospitals within the MSHS in New York City: the Mount Sinai Hospital (MSH) located in East Harlem, Manhattan; Mount Sinai Morningside (MSM) located in Morningside Heights, Manhattan; Mount Sinai West (MSW) located in Midtown West, Manhattan; Mount Sinai Brooklyn (MSB) located in Midwood, Brooklyn; and Mount Sinai Queens (MSQ) located in Astoria, Queens. The data set was obtained from different sources using the Epic EHR software (Epic Systems) and aggregated by the Mount Sinai COVID Informatics Center.

Study Population

We retrospectively included all patients who were over 18 years of age, had laboratory-confirmed COVID-19 infection, and were admitted to any of the abovementioned MSHS hospitals between March 15 and May 22, 2020. A confirmed case of COVID-19 was defined by a positive reverse transcriptase–polymerase chain reaction (RT-PCR) assay of a nasopharyngeal swab. To restrict our data to only primary COVID-19–related encounters, we excluded patients who had a first positive COVID-19 RT-PCR result more than two days after admission. We included all patients who had been discharged, had died, or were still admitted and had stayed in the hospital for at least the amount of time corresponding to the outcome in question. This approach provided additional training data for the initial timeframes described in the paper. All exclusion criteria are presented in Figure 1A.

Figure 1. Study design and workflow. (A) Procedure for patient inclusion in our study. (B) Outcomes of interest. We trained the model on data taken at time of admission to predict the likelihood of either mortality or critical event occurrence at 3, 5, 7, and 10 days. (C) Strategy and design of the experiments. Patient clinical data from Mount Sinai Hospital (MSH) before the temporal split (May 1) were used to train and internally validate our XGBoost model in comparison with other baseline models. We then tested the series of XGBoost models on unimputed patient data on patients from four other external hospitals within the MSHS for external validation. h: hours; ICU: intensive care unit; lab: laboratory; MSB: Mount Sinai Brooklyn; MSHS: Mount Sinai Health System; MSM: Mount Sinai Morningside; MSQ: Mount Sinai Queens; MSW: Mount Sinai West (MSW); RT-PCR: reverse transcriptase–polymerase chain reaction; vitals: vital signs.
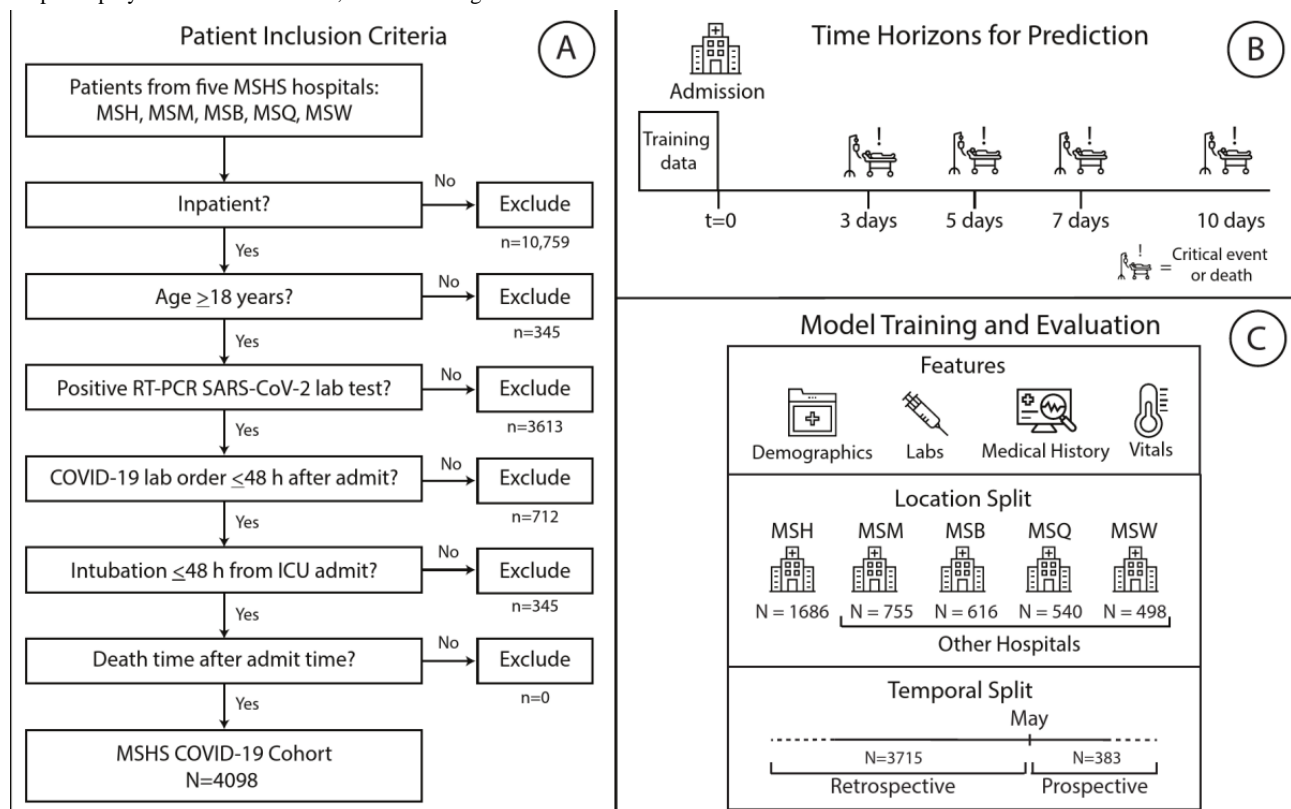
Study Design

We built predictive models based on data from MSH patients who were admitted from March 15 to May 1, 2020, which was the cutoff time for prospective evaluation. These patients were considered to be part of the internal validation cohort. All patients admitted to other hospitals (OH) were grouped together. To allay concerns about effects of immortal time bias and

censoring on the results, we recorded the ultimate outcome of each patient who was admitted in this time frame, even if the outcome occurred after the data enrollment cutoff. For patients within the internal validation cohort, the models were trained and their performance was evaluated through stratified k-fold cross-validation to mitigate the variability of a single train-test split. A final model was then trained for each outcome and time window using all the patients in this data set, and this model

was then assessed through a series of validation experiments. First, we externally validated OH patients from March 15 to May 1, 2020, which was the same time frame used to train the model; this afforded benefits by assessing the generalizability of the model to a new setting (Figure 1B). Then, to assess temporal generalizability, we performed prospective validations of the model independently on both MSH and OH patients admitted from May 1 to May 22, 2020 (Figure 1C).

Study Data

Demographics collected included age, sex, reported race, and ethnicity. Race was collapsed into seven categories based on the most recent US census race categories: American Indian or Alaskan Native, Asian, Black or African American, other, Native Hawaiian or other Pacific Islander, unknown, and White [23]. Ethnicity was collapsed into three categories: Hispanic/Latino, non-Hispanic/Latino, and Unknown.

Additionally, diagnosis codes based on International Classification of Diseases-9/10-Clinical Modification (ICD-9/10-CM) codes and procedures were obtained to identify associated pre-existing conditions. We chose to include conditions with previously reported increased incidence in hospitalized patients with COVID-19: coronary artery disease, heart failure, hypertension, atrial fibrillation, obstructive sleep apnea, asthma, chronic obstructive pulmonary disease, cancer, chronic kidney disease, diabetes, viral hepatitis, liver disease, intracerebral hemorrhage, and stroke [9,24-27]. Inclusion of these chronic conditions and acute kidney injury (AKI) was based on ICD-9/ICD-10 codes related to active problems documented during COVID-19 hospitalization, defined by the presence of at least one ICD code signifying the condition. Laboratory measurements and vital signs near the time of admission were also retrieved for each patient during their hospital encounter. Given the resource constraints due to COVID-19, which delayed acquisition of laboratory results, the first laboratory value in a 36-hour window period was used as the representative laboratory value on admission. The implications of this strategy for the model performance are illustrated in the Multimedia Appendices.

All laboratory orders from the five hospitals were queried for patients included in this study within the timeframe of interest. Due to discrepancies in how laboratory orders were named in different hospitals, a comprehensive and statistical review of all laboratory orders by field name was conducted by a multidisciplinary team of clinicians to ensure direct mapping between all sites. Additionally, many laboratory values represented a single component (eg, sodium) but were acquired from either an arterial blood gas (ABG) test, venous blood gas (VBG) test, or basic metabolic panel (BMP). Based on the utility of these laboratory values in clinical practice and the similarity between their statistical distributions, laboratory values derived from a VBG or BMP were collapsed into a single category (ie, "SODIUM") and those derived from an ABG were moved to a separate category (ie, "SODIUM_A"). In the set of all laboratory order names that were combined into a single laboratory category, the earliest laboratory result by time was chosen as the representative laboratory value for that category. Finally, laboratory data below the 0.5th percentile and above the 99.5th percentile were removed to avoid inclusion of any obvious outliers that could represent incorrect documentation or measurement error.

Data Sharing

The raw data used in this work cannot be shared due to patient privacy and security concerns. However, we are open to using this data set for validation of other models through a collaboration under an appropriate data use agreement with the authors at the Icahn School of Medicine at Mount Sinai.

Definition of Outcomes

The two primary outcomes were (1) death versus survival or discharge and (2) critical illness versus survival or discharge through time horizons of 3, 5, 7, and 10 days. Critical illness was defined as discharge to hospice, intubation ≤48 hours prior to intensive care unit (ICU) admission, ICU admission, or death. A composite outcome (ie, mortality as opposed to discharge or survival) was chosen to bypass issues of competing risks.

Model Development, Selection, and Experimentation

Our primary model was the Extreme Gradient Boosting (XGBoost) implementation of boosted decision trees on continuous and one-hot encoded categorical features [28]. The XGBoost algorithm provides robust prediction results through an iterative process of prediction summation in decision trees fit to the residual error of the prior ensemble. While each tree is too simple to accurately capture complex phenomena, the combination of many trees in the XGBoost model accommodates nonlinearity and interactions between predictors. The model directs missing values through split points to minimize loss. Hyperparameter tuning was performed by randomized grid searching directed toward maximizing the F1 score metric over 5000 discrete grid options. Ten-fold stratified cross-validation was performed inside each grid option, and the optimal hyperparameter set was chosen based on the model in the grid search with the highest F1 score. Final model hyperparameters for the XGBoost model are listed in Multimedia Appendix 1. To generate confidence intervals for the internal validation set, training and testing was performed for 500 bootstrap iterations with a unique randomly generated seed for the train-test data splits.

We opted to implement our analyses within a classification framework because we aimed to implement our models with regard to clinically relevant time boundaries for resource allocation and clinical decision-making, such as resource allocation, triage, and decisions for ICU transfer. A major goal of our analysis was the construction of a resilient and highly performant predictive model; therefore, the selection of the XGBoost algorithm is reasonable given its well-understood properties as the best-performing machine learning algorithm for classification tasks on tabular data. The XGBoost algorithm also addresses real-life problems such as missing data and highly multidimensional independent variables, while alternate strategies and extensions must be employed to enable Cox proportional hazard analyses in these settings.

To compare the performance of our XGBoost model for the training and internal validation data, we generated two predictive

models as a baseline, namely logistic regression (LR) and LR with L1 regularization, given their ubiquity as preferred models in current COVID-19 research. L1 regularization, also known as least absolute shrinkage and selection operator (LASSO), was used to train the LR and impose parsimony in feature selection, given the number of features present in the data set (73). LASSO and LR were optimized by an exhaustive grid search for the inverse regularization parameter (Multimedia Appendix 1). For these baseline models, the issue of missingness was addressed by imputation. Features with >30% missingness were dropped, and k-nearest neighbors (kNN, k=5) was used to impute missing data in the remaining feature space. To further assess the impact of imputation on performance, an XGBoost model was also created and trained on the imputed data set. Imputation for the training set (ie, MSH only) and external validation set (ie, OH) were performed using only the first collected value from the respective sites to prevent information leakage that could compromise assessment of generalizability. We assessed the calibration of the results of each model to ensure that the model predictions could be interpreted as real-world risk scores. Calibration was performed using both the sigmoid and isotonic methods of the *CalibratedClassifierCV* class in scikit-learn and evaluated using the Brier score metric.

Experimental Evaluation

All models were trained and evaluated using 10-fold stratified cross-validation, and confidence intervals were generated using 500 iterations of bootstrapping. Stratified k-fold cross-validation maintains an outcome distribution across each fold in concordance with the outcome distribution in the study population. We present calibration plots for all these experiments, including isotonic and sigmoid calibrations, that show the proportion of positive cases to the mean predicted value for the raw models in Figures S1-S8 in Multimedia Appendix 2. In these plots and in Multimedia Appendix 3, we also report the Brier score, which measures the quality of calibration (a lower score indicates greater accuracy). Ultimately, we selected the best-calibrated model based on the lowest Brier score, and performed all subsequent experiments with this model. Probability scores output by the model were used to calculate the areas under the receiver operator characteristic curve (AUC-ROCs) and areas under the precision-recall curve (AU-PRCs). The receiver operating characteristic curve shows how the balance between true and false positive rates is affected at different decision thresholds. The precision-recall curve visualizes how the balance of false positives and negatives is affected at different decision thresholds. The decision threshold was calculated separately for each fold to maximize the F1 score for prediction of the primary outcome. The threshold for the final model was taken as the median of the calculated thresholds across the 10 cross-validation folds. Accuracy, F1 score, sensitivity, and specificity were calculated on the basis of these thresholds. Model performance was assessed during internal cross-validation, external validation, and prospective validation. The models were compared on the basis of their AUC-ROC and AU-PRC values across the time intervals in each population of patients. The AU-PRC is known to be a better metric in skewed data sets that have greater class imbalance and was therefore primarily used in the model evaluation and selection.

Model Interpretation

We evaluated feature contributions toward model prediction using SHAP scores. SHAP scores are a game-theoretic approach to model interpretability; they provide explanations of global model structures based upon combinations of several local explanations for each prediction [29]. To interpret and rank the significance of input features toward the final prediction of the model, mean absolute SHAP values were calculated for each feature across all observations in the internal validation set. We also plotted a heatmap showing SHAP interaction values, which are an extension of SHAP summary values to capture how pairwise interactions between different features contribute to model prediction. For a given pair of features, their pairwise interaction effect is calculated after removing the individual effects of those features. Values on the diagonal represent the main effects (ie, the SHAP summary values), and values off the diagonal represent the interaction effects. Higher values on the heatmap (ie, brighter squares) represent a greater impact on model predictions. In addition, we calculated the feature importance from the coefficients of the LASSO model (Figure S9 in Multimedia Appendix 2).

Promoting Interoperability and Replicability

This article is written following the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines [30], which are further elaborated in Multimedia Appendix 4. Furthermore, we release all code used to build the classifier under the GPLv3 license in a public GitHub repository [31].

Results

Clinical Data Source and Study Population

Electronic health records for 4098 COVID-19–positive inpatient admissions at five hospitals within the MSHS between March 15 and May 22, 2020, were retrieved for data analysis based on the inclusion criteria. These data included patient demographics, past medical history, and admission vital signs and laboratory test results (Table 1 and Table 2; Multimedia Appendix 5). Vital sign and laboratory test data were included as baseline features in order to work within the bounds of the processing and operations involved in obtaining the results of these tests. No data leakage occurred, and we did not find disproportionate rates of feature missingness for patients who died within this time window for feature inclusion (see the Multimedia Appendices). We show the number of patients involved and the proportion of events in each experiment by time window in Multimedia Appendix 6. Relevant patient events (intubation, discharge to hospice care, or death) were recorded, and subsets were constructed at 3-, 5-, 7-, and 10-day intervals after admission (Figure 1). Before May 1, 21.3% to 35.3% of patients had experienced a critical event (intubation, ICU admission, discharge to hospice care, or death) across all time intervals. On or after May 1, this proportion changed to 14.3% to 21.9%. Similarly, before May 1, 2.6% to 22.4% patients died across all time intervals, with the proportion changing to 1.1% to 8.0%

on or after May 1. The survival curve for mortality is shown in Figure S10 in Multimedia Appendix 2. This curve was generated by fitting a Kaplan-Meier estimator to the survival time for patients with observed (in-hospital) death instead of discharge (Multimedia Appendix 6). In contrast, the set of noncases consisted of patients with all other discharge dispositions and those who were still hospitalized at the respective intervals after admission.

Table 1. Demographic characteristics, clinical history, and vital signs of hospitalized patients with COVID-19 at baseline (N=4098).

Characteristic on admission	Retrospective		Prospective	
	MSH[a] (n=1514)	OH[b] (n=2201)	MSH	OH
Demographics				
Sex, n (%)				
Male	869 (57.4)	1257 (57.1)	104 (59.4)	104 (50)
Female	645 (42.6)	944 (42.9)	71 (40.6)	104 (50)
Race, n (%)				
Other	639 (42.2)	804 (36.5)	80 (45.7)	53 (25.5)
Caucasian	354 (23.4)	533 (24.2)	43 (24.6)	56 (26.9)
African American	357 (23.6)	688 (31.3)	37 (21.1)	79 (38)
Unknown	80 (5.3)	45 (2)	__c	—
Asian	77 (5.1)	102 (4.6)	10 (5.7)	11 (5.3)
Pacific Islander	—	—	—	—
Ethnicity, n (%)				
Non-Hispanic/Latino	820 (54.2)	1377 (62.6)	98 (56)	139 (66.8)
Hispanic/Latino	421 (27.8)	556 (25.3)	50 (28.6)	43 (20.7)
Unknown	271 (17.9)	236 (10.7)	24 (13.7)	26 (12.5)
Age, median (IQR)	62.9 (50.7-73)	69.6 (53.3-80)	63.7 (51.2-73.8)	69.8 (55.5-79.9)
Age (years), n (%)				
18-30	64 (4.2)	46 (2.1)	16 (9.1)	—
31-40	155 (10.2)	113 (5.1)	13 (7.4)	12 (5.8)
41-50	165 (10.9)	160 (7.3)	14 (8)	17 (8.2)
51-60	291 (19.2)	341 (15.5)	33 (18.9)	35 (16.8)
61-70	394 (30)	517 (20)	40 (20)	39 (20)
71-80	258 (17)	522 (23.7)	41 (23.4)	52 (25)
81-90	142 (9.4)	396 (18)	13 (7.4)	38 (18.3)
≥90	45 (3)	106 (5)	—	—
Previous medical history, n (%)				
Hypertension	64 (4.2)	46 (2.1)	63 (40)	83 (40)
Atrial fibrillation	155 (10.2)	113 (5.1)	13 (7)	21 (10)
Coronary artery disease	165 (10.9)	160 (7.3)	32 (20)	41 (20)
Heart failure	291 (19.2)	341 (15.5)	26 (10)	30 (10)
Stroke	394 (30)	517 (20)	16 (9)	10 (5)
Chronic kidney disease	258 (17)	522 (23.7)	32 (20)	43 (20)
Diabetes	142 (9.4)	396 (18)	40 (20)	54 (30)
Asthma	45 (3)	106 (5)	11 (6)	—
Chronic obstructive pulmonary disease	64 (4.2)	46 (2.1)	13 (7)	11 (5)
Cancer	158 (10)	124 (6)	43 (20)	14 (7)
Vital signs at hospital admission, median (IQR)				
Heart rate (beats per minute)	87 (77-97)	86 (76-98)	85 (74-97.5)	82 (72.8-96)
Pulse oximetry (%)	96 (94-97)	96 (94-98)	97 (95-98)	97 (96-98)
Respiration Rate (breaths per minute)	20 (18-21)	18 (18-20)	18 (18-20)	18 (18-20)

Characteristic on admission	Retrospective		Prospective	
	MSH[a] (n=1514)	OH[b] (n=2201)	MSH	OH
Temperature (ºF)	98.7 (98-99.9)	98.5 (97.7- 99.3)	98.1 (97.5-98.6)	97.9 (97.3-98.6)
Systolic blood pressure (mm Hg)	125 (112-140)	125 (111-140)	122 (111.5-138)	127 (112.8-141.2)
Diastolic blood pressure (mm Hg)	69 (61-78)	72 (64-80)	70 (60.5-78.5)	72 (64-82)
BMI (kg/m^2)	28.1 (24.4-32.8)	27.5 (24.2-32.5)	25.92 (21.9-30.4)	27.7 (23.4-32.1)

[a]MSH: Mount Sinai Hospital.

[b]OH: other hospitals.

[c]—: Values with fewer than 10 patients per field are censored to protect patient privacy.

Table 2. Admission laboratory parameters of hospitalized patients with COVID-19 at baseline (N=4098), median (IQR).

Laboratory parameters	Retrospective		Prospective	
	MSH[a] (n=1514)	OH[b] (n=2201)	MSH	OH
Metabolic markers				
Sodium (mEq/L)	138 (135-140)	139 (136-142)	139 (136-141)	139 (136-141)
Potassium (mEq/L)	4 (3.6-4.5)	4.3 (3.9-4.7)	4 (3.7-4.4)	4.3 (3.8-4.6)
Creatinine (mg/dL)	0.91 (0.7-1.5)	1.01 (0.8-1.7)	0.89 (0.7-1.6)	1.12 (0.7-2.1)
Lactate (mg/dL)	1.8 (1.4-2.3)	1.4 (1.1-2)	1.8 (1.4-2.3)	1.49 (1-1.9)
Hematological markers				
White blood cells (10^3/μL)	7 (5-10.2)	7.6 (5.5-10.9)	7.3 (5.1-10.7)	8.3 (6.3-11.9)
Lymphocyte percentage	NA (NA-NA)	14.2 (8.6-21.3)	NA (NA-NA)	14.7 (9.9-21.6)
Hemoglobin (mEq/L)	12.2 (10.7-13.5)	12.7 (11.1-13.9)	10.5 (9.1-12.8)	11.1 (9.2-12.8)
Red blood cell distribution width (%)	4.2 (3.7-4.6)	4.28 (3.8-4.7)	3.69 (3.1-4.3)	3.79 (3.2-4.5)
Platelets (n)	220 (165-291)	208 (158-281)	224 (166.2-304)	211 (149.2-285.2)
Liver function				
Alanine aminotransferase (units/L)	30 (18-53)	31 (19-54)	26 (13.8-51)	23 (14-36)
Aspartate aminotransferase (units/L)	42 (28-66)	45 (30-74)	30 (20-50.5)	30 (19-49)
Albumin (g/dL)	2.9 (2.5-3.2)	2.9 (2.5-3.2)	2.9 (2.5-3.4)	2.9 (2.3-3.3)
Total bilirubin (mg/dL)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.7 (0.4-1)	0.5 (0.4-0.7)
Coagulation markers				
Prothrombin time (s)	14.5 (13.6-16)	14.9 (13.9-16.5)	14.8 (13.6-16.2)	15.05 (13.7-17.6)
Partial Thromboplastin time (s)	32.9 (29.2-38.5)	34.8 (30.3-41.5)	32.6 (28.8-37.8)	36.1 (31-45.9)
Gases				
PCO_2[c] (mmHg)	42 (37-47)	42 (37-53)	44 (39-49)	42 (37-48.5)
pH	7.4 (7.3-7.4)	7.36 (7.3-7.4)	7.39 (7.4-7.4)	7.36 (7.3-7.4)
Inflammatory markers				
C-reactive protein (mg/L)	116.4 (57.1-199.5)	132.4 (65.8-218.9)	62.2 (17-148.9)	73.7 (33.5-181.8)
Ferritin (ng/mL)	800 (365-1916)	906 (438-2056)	485 (200.2-1031.5)	690 (303.5-1470.2)
D-dimer (ng/mL)	1.44 (0.8-3)	2.42 (1.2-4.4)	1.66 (0.9-3.1)	1.97 (1.1-3.8)
Creatinine phosphokinase (units/L)	146 (70-488)	220 (76.8-501.8)	194.5 (93.2-290.8)	271.5 (48.8-611.5)
Lactate dehydrogenase (units/L)	423 (315-571)	466.5 (356.2-652.2)	334 (251.5-472)	364 (266.8-487)
Cardiac markers				
Troponin I (ng/mL)	0.05 (0-0.2)	0.064 (0-0.2)	0.05 (0-0.1)	0.0525 (0-0.1)

[a]MSH: Mount Sinai Hospital.

[b]OH: other hospitals.

[c]PCO_2: partial pressure of carbon dioxide.

Classifier Training and Performance

We developed models based on cross-validation experiments for all model types and conditions within the MSH at the earlier time period of the study (ie, up to the enrollment date cutoff). On internal cross-validation for mortality prediction, the unimputed XGBoost model demonstrated strong performance, with AUC-ROC values ranging from 0.84 to 0.90 and AU-PRC values ranging from 0.33 to 0.48 (Multimedia Appendix 6). In comparison, LR and LASSO, after kNN imputation on the

missing data elements, performed marginally worse on every outcome, with AUC-ROC values ranging from 0.80 to 0.82 and AU-PRC values ranging from 0.10 to 0.40 (Figure 2, Multimedia Appendix 6). Additionally, when trained with imputed data, the XGBoost classifier performed worse, achieving AUC-ROC values from 0.80 to 0.84 and AU-PRC values from 0.18 to 0.40 across all time periods. In the case of internal validation for critical event prediction, the AUC-ROC values of the unimputed XGBoost model ranged from 0.79 to 0.81, and the AU-PRC values ranged from 0.60 to 0.70. The performance for the

LASSO and LR models with imputation was poorer, with AUC-ROC values of 0.75 to 0.77 and AU-PRC values of 0.54 to 0.65.

We then assessed the performance of these models in three validation experiments: in OH within the same time period, within the same hospital (MSH) at a future time period, and in OH at a future time period. The unimputed XGBoost AUC-ROC generally showed the best performance for mortality prediction across intervals ranging from 0.84 to 0.88, with AU-PRC values ranging from 0.44 to 0.64. For LR and LASSO, the AUC-ROC values ranged from 0.82 to 0.83, while the AU-PRC values ranged from 0.22 to 0.58. The imputed XGBoost continued to perform slightly worse, with AUC-ROC values ranging from 0.72 to 0.83 and AU-PRC values ranging from 0.17 to 0.60. For critical event prediction across all time intervals, the AUC-ROC values of the unimputed XGBoost model ranged from 0.78 to 0.81, while the AU-PRC values ranged from 0.51 to 0.69. Performance of LR and LASSO was marginally worse, with ranges of 0.74 to 0.81 for the AUC-ROC and 0.44 to 0.70 for the AU-PRC. The performance of imputed XGBoost was similar to that of unimputed XGBoost, with AUC-ROCs ranging from 0.76 to 0.82 and AU-PRCs ranging from 0.49 to 0.71.

Similarly, in prospective validation at OH for mortality prediction across all time intervals, the AUC-ROC values of the unimputed XGBoost model ranged from 0.68 to 0.88, while the AU-PRC values ranged from 0.13 to 0.31. The performance of the LR and LASSO models in the same experiments was much poorer in terms of AUC-ROC, with values ranging from 0.51 to 0.74, and at par with the unimputed XGBoost model in terms of AU-PRC, with values ranging from 0.13 to 0.34. Imputed XGBoost performed worse overall, with AUC-ROC values ranging from 0.66 to 0.81 and AU-PRC values between 0.06 and 0.21. In the case of prospective validation for critical event prediction at OH, the AUC-ROC values of the unimputed XGBoost model ranged from 0.74 to 0.77, and the AU-PRC

values were between 0.36 and 0.50. In contrast, the performance of the LR and LASSO models over the same conditions was poorer overall, with ranges of 0.65 to 0.74 for the AUC-ROC and 0.31 to 0.46 for the AU-PRC. The imputed XGBoost model again performed slightly worse than the unimputed XGBoost model, with AUC-ROC values from 0.71 to 0.77 and AU-PRC values between 0.31 and 0.48. Prospective validation at MSH presented a new set of challenges for all the models because of the generally lower number of outcomes and larger class imbalance for mortality prediction for the shorter time intervals. For mortality prediction overall, the AUC-ROC values of the unimputed XGBoost model ranged from 0.85 to 0.96, and the AU-PRC values ranged from 0.32 to 0.55. The LR and LASSO models showed much poorer performance, with AUC-ROC values ranging from 0.44 to 0.85 and AU-PRC values ranging from 0.01 to 0.41. The imputed XGBoost model also performed worse than the unimputed XGBoost model, with AUC-ROC values of 0.82 to 0.88 and AU-PRC values of 0.04 to 0.50. For prediction of critical events, the AUC-ROC values of the unimputed XGBoost model were between 0.72 and 0.78, and its AU-PRC values were between 0.40 and 0.54. The performance of the LR and LASSO models in the same set of experiments was slightly poorer, with ranges of 0.66 to 0.75 for the AUC-ROC and 0.32 to 0.48 for the AU-PRC. The imputed XGBoost model performed marginally worse than the unimputed XGBoost model, with values of 0.71 to 0.77 for the AUC-ROC and 0.42 to 0.50 for the AU-PRC.

Model calibration as measured by Brier scores improved after either sigmoid or isotonic calibration across all time windows. For the unimputed XGBoost model, isotonic calibration performed better than sigmoid calibration, with Brier scores ranging from 0.124 to 0.161 for critical event prediction and from 0.019 to 0.085 for mortality prediction. Sigmoid calibration only slightly outperformed isotonic calibration for critical event prediction at 10 days (Brier scores of 0.160 vs 0.161, respectively).

Figure 2. Comparison of the performance of the XGBoost and baseline models. Performance of the XGBoost classifier by ROC curves (left) and PR curves (right) on the unimputed data set (red) for mortality (top) and critical event (bottom) prediction versus the three baseline models: XGBoost classifier on the imputed data set (purple), LASSO (green), and LR (orange). LASSO: least absolute shrinkage and selection operator; PRC: precision-recall curve; ROC: receiver operating characteristic; XGB: Extreme Gradient Boosting.
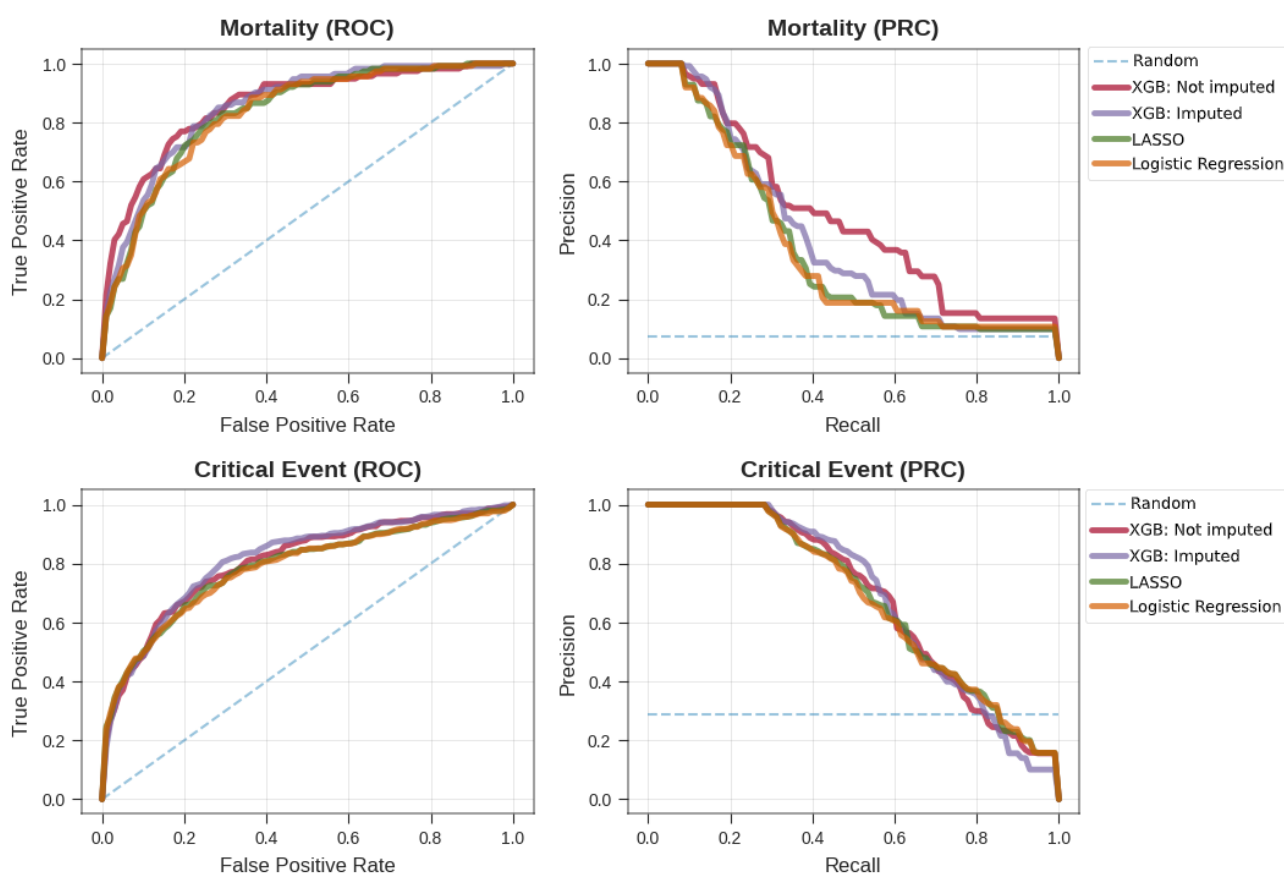
Model Feature Importance

Mean absolute SHAP values [32] were calculated for each XGBoost model in the internal validation data set (Figure 3). For critical event prediction, the presence of acute kidney injury and both high and low levels of lactate dehydrogenase (LDH), respiratory rate, and glucose were strong drivers for predicting a critical event within one week. Other notable drivers of predictability included both systolic and diastolic blood pressure, pH, total protein levels, C-reactive protein, and D-dimer. For mortality, both high and low values for age, anion gap, C-reactive protein, and LDH were the strongest effectors in guiding mortality prediction within one week of admission. Other important variables for increasing the prediction of death included oxygen saturation on intake admission, blood urea nitrogen, ferritin, red cell distribution width (RDW), diastolic blood pressure, and lactate. Finally, using SHAP interaction scores, we discovered that covariate interactions between features contributed less to the predictions of the models than the independent importance of each feature (Figures S11 and S12 in Multimedia Appendix 2), except for the case of AKI,

where levels of LDH, glucose, and C-reactive protein were strong covariates. As a comparison, we also assessed the feature importance for the LASSO model for these experiments (Figure S9 in Multimedia Appendix 2). We saw an overlap of key features that both models considered important in their predictions for both critical event and mortality prediction at 7 days. For critical events, we found that AKI was the most important feature in both models. Higher respirations and D-dimer levels were also associated with higher mortality, and lower diastolic blood pressure was negatively associated. For mortality, we also saw strong concordance in key features between both models. Specifically, older age and higher anion gap were strong contributors to mortality prediction in both models, and lower diastolic blood pressure and oxygen saturation were negatively associated with mortality. It is encouraging that many of the features with high importance in the primary XGBoost model were also prioritized in the LASSO classifier, suggesting the robustness of the predictive ability of these features. The top 10 features for the Critical Event and Mortality models at seven days are enumerated in Multimedia Appendix 7.

Figure 3. Performance of the XGBoost classifier by ROC curves (left) and precision-recall curves (right) for mortality (top) and critical events (bottom) in validation experiments of generalizability and time. For generalizability, we show our XGBoost model from cross-validation on MSH and applied to all other hospitals. We also show the performance of the model on prospective patients who were unseen at the time of the original experiment at MSH and all other hospitals in the same time frame. Ext. Val.: external validation; Int. Val.: internal validation; MSH: Mount Sinai Hospital; OH: other hospitals; PRC: precision-recall curve; Prosp. Val: prospective validation; ROC: receiver operating characteristic.
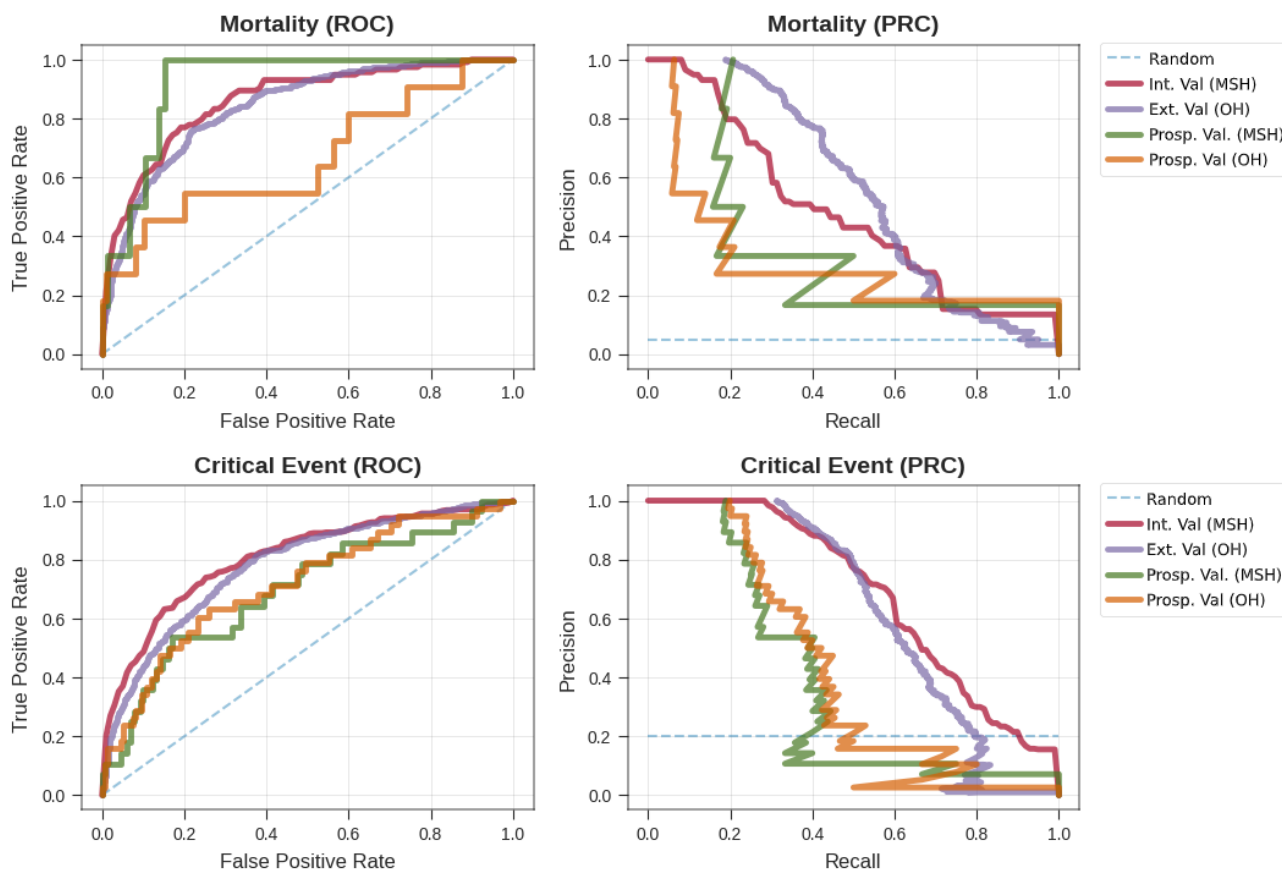
Discussion

Principal Findings

In this work, we performed a series of experiments with the goal of using machine learning to predict in-hospital mortality or critical events from admission for patients with COVID-19. We highlight several important findings with implications for clinical medicine. First, we offer a robust prediction algorithm pertaining to the most clinically severe outcomes based solely on admission metrics, which maintains its training performance in both external and prospective validation experiments. Most notably, the high specificity in predicting mortality within 3, 5, and 7 days of admission (AU-PRCs of 0.91 to 0.97) suggests a role of the algorithm in augmenting clinicians' decision-making when identifying patients at immediate risk of impending clinical decompensation and potential in guiding allocation of more intensive care upon admission. Finally, the impact of the large class imbalance and missingness on model training and performance can be appreciated when comparing mortality predictions at 3 days. On the non-imputed data set, the XGBoost classifier achieves a remarkably higher AU-PRC (0.44) compared to the models using imputed data (0.14 for LR and LASSO, 0.12 for XGBoost imputed). It is important to note the consideration of the AU-PRC instead of the AUC-ROC for

deriving this claim, as the AU-PRC includes both precision (ie, positive predictive value) and recall (ie, sensitivity) and thus accounts for the class imbalance, which the AUC-ROC metric generally ignores. Overall, we found that the unimputed XGBoost model performed better not only in internal validation but in the vast majority of the other validation experiments. As such, we believe it can be generalized more readily than the other models to new cohorts and time points. Along these lines, we found that our imputation strategy generally hindered the performance of the XGBoost model. There were instances where the XGBoost model performed approximately the same (within the bounds of the confidence intervals) or worse than the other comparators for different metrics. For instance, in the prospective OH experiment for predicting critical events within 7 and 10 days, the LASSO method outperformed the unimputed XGBoost model in terms of AUC-ROC and AU-PRC. In the 7-day condition, however, the imputed XGBoost model actually performed the best overall, which suggests that the imputation strategy worked better in this particular scenario. Additionally, in the prospective OH experiment, the unimputed XGBoost model underperformed compared to the other models for mortality prediction; however, we believe this was due to the extremely low positive prevalence. Thus, while XGBoost makes assumptions on how it handles missing data, we found that

XGBoost without imputation was the more robust method in these experiments. Furthermore, this strategy is conducive to implementation into clinical operations, as it removes the need for an intermediary imputation step.

Additionally, our framework permits a clinically relevant understanding of the most salient features of the unimputed XGBoost model, defining its decision boundaries using patients from the holdout set during internal validation (Figure 4). At 7 days, age was the most important feature for mortality prediction in COVID-19–positive patients, with a notably rapid and nonlinear increase of feature contribution with increasing age (Figure 4) [33,34]. Hyperglycemia, particularly in the ranges that catered to positive predictions (Figure 4C), may serve as proxies for either metabolic syndrome, diabetic ketoacidosis, or hyperosmolar hyperglycemic state predisposition from underlying diabetes, which have previously also been reported and associated with poorer outcomes in COVID-19–positive patients [35-37]. The higher information content in continuous values such as glucose levels and their larger role in the level of control of diabetes is a likely explanation for why diabetes, as a comorbidity, failed to be a strong driver of prediction. The demonstration of the anion gap, in conjunction with high levels of lactate, as another strong model influencer for mortality prediction is likely linked with potential ongoing elevated anion-gap metabolic acidosis from a brewing severe inflammatory response syndrome or sepsis picture [38]. Elevation in serum LDH is a nonspecific marker of inflammation; however, it is implicated in pulmonary endothelial cell injury and in COVID-19–positive patients [39-41]. AKI has been reported in patients with severe COVID-19 and, if present early, may be a strong indicator of future critical events [42,43]. The covariate relationship between LDH, CRP, and glucose may reflect underlying severe inflammation and deranged metabolism, which may be contributing to the AKI. Elevated RDW, which may be an index of enhanced patient frailty and risk of adverse outcomes [35], was also a strong driver of mortality. Additionally, vital sign instability (low oxygen saturation, tachypnea, hypotension), elevated ferritin [41,44], high lactate, and acidosis were contributors to driving model predictions toward mortality. With growing evidence of COVID-19–induced hypercoagulable states in these patients [41,45,46], it is promising that our model recognized the feature importance of coagulability markers such as D-dimer (Figure 4). Thus, this corroboration of the features learned by XGBoost and highlighted by the SHAP analysis with the findings from pathophysiological principles and more recent correlative studies exploring patients with COVID-19 [2,3,9,25,26,47,48] gives additional credibility to these findings. Additionally, when we compared these features to those that were ranked highly for the LASSO model, we found many concordant features with the same direction of effect; this further strengthens the evidence of the utility of these features in predictive models (Figure S9 in Multimedia Appendix 2).

Just as interesting as the most important features identified for classification by XGBoost are the features that were not prioritized (ie, much lower mean absolute SHAP values). For example, race is a social construct that at best serves as a proxy for the social disparities leading to infection risk at the population level, and it is also related to the distribution of comorbid conditions that potentiate disease severity. Furthermore, race is both poorly represented (including a category for "Unknown") and inadequately characterized in the EHR. While race, in and of itself, potentially carries a large amount of information because it inadvertently represents the very societal inequities that lead to poorer outcomes (ie, structural racism as a contributor to COVID-19 health disparities), the model instead chose to prioritize more objective markers of health status (laboratory values, vital signs, comorbidities) that more directly represent the deeper biology of the risk factors and state of disease severity leading to these adverse outcomes. Contrary to our expectation, age was not identified as a significant feature for critical event prediction within 7 days in the primary analyses. This suggests that the model decided to capture acute critical events by relying on more objective measures that are not confounded by other factors that are cached into age, which may better represent illness severity and more irreversible outcomes (ie, death). Age may then be a better marker for mortality by offering a more stable container of clinical information, given its invariance to change relative to other features.

Figure 4. SHAP summary plots for critical event (A) and mortality (D) at 7 days showing the SHAP values for the 10 most important features for the respective XGBoost models. Features in the summary plots (y-axis) are organized by their mean absolute SHAP values (x-axis), which represent the importance of the features in driving the prediction of the classifiers for patients. (B) and (C) Dependency plots demonstrating how different values can affect the SHAP score and ultimately impact classifier decisions for LDH and glucose, respectively, for critical event prediction. (E) and (F) Dependency plots for age and C-reactive protein levels. Patients with missing values for a feature in the dependency plot are clustered in the shaded area to the left. LDH: lactate dehydrogenase; RDW: red cell distribution width; SHAP: SHapley Additive exPlanation.
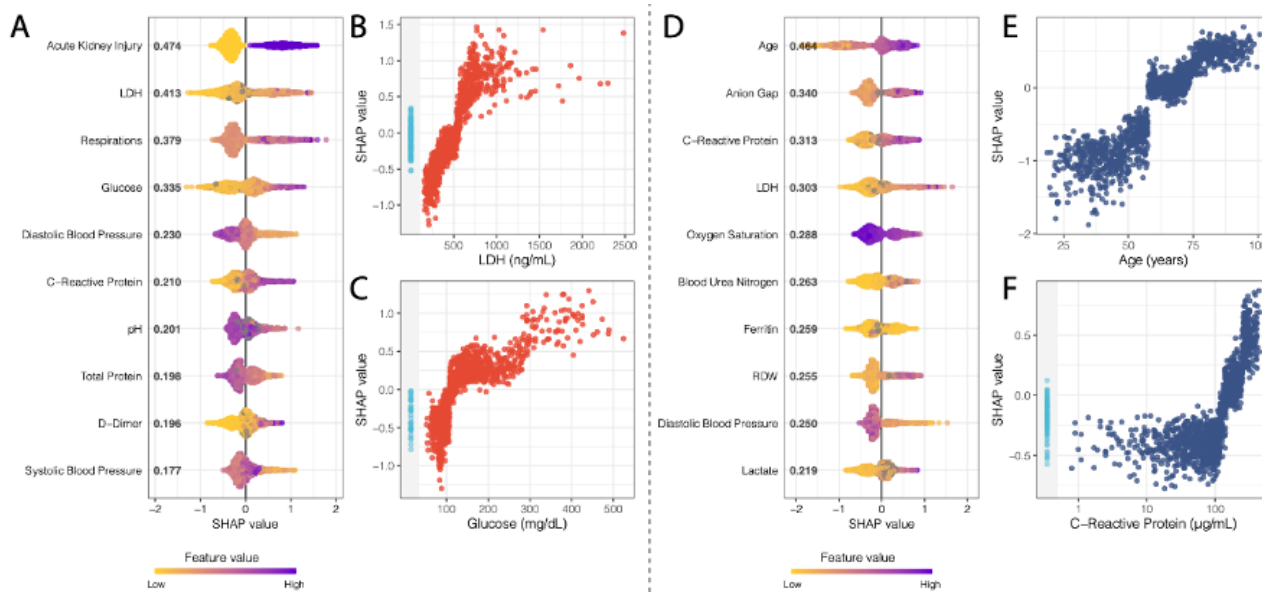
Limitations

The results of our models should be considered in light of several limitations. First, we based our predictions solely on data extracted around patient admission (ie, within 36 hours). This step was added purposefully to remove potential bias from effects of hospital workflow, and we found that it did not cause another source bias relating to informed missingness (see Multimedia Appendices). No information from the future was leaked into this prediction. Although the restriction of using data at admission encourages the use of this model in patient triage, events during a patient's hospital stay after admission may drive their clinical course away from the prior probability, which cannot be captured by baseline admission features. We believe a "live" or continuously updating modelling approach would be better suited for this as a future direction. Furthermore, not all patient laboratory values are drawn at admission, which introduces an element of missingness in our data set. For example, unlike the general patient population, patients on anticoagulation therapy, who likely have comorbidities increasing their baseline risk, will have coagulation laboratory tests (prothrombin time, partial thromboplastin time) performed on admission. We attempted to mediate this issue by including a missingness threshold cutoff, assessing model performance with imputation, and not including any laboratory test that was specific to an intervention (ie, arterial laboratory tests performed in the ICU). Additionally, patients admitted to the hospital later in the crisis benefited from improved patient care protocols from experiential learning but were also negatively affected by resource constraints from overburdened hospitals. These effects may also induce temporal variation between patient outcomes, which is demonstrated by the lower critical event and mortality rate in the prospective validation data set. However, determining the models' performance in this scenario was one of the

justifications for including a future time point. Despite a certain dip in overall performance for the unimputed XGBoost model, which we attribute to heavy imbalance of outcomes and extremely low prevalence rates, we were overall encouraged by its performance. Furthermore, inherent limitations exist when using EHRs, especially those integrated from multiple hospitals. To facilitate timely dissemination of our results, we chose not to manually chart review patient notes that may have otherwise provided additional potential features, such as symptoms and clinical course, to incorporate in our model. Because all five hospitals operate in a single health system, system-wide protocols in laboratory order sets and management protocols were an additional source of bias that may lower external validity. Other interhospital effects, such as shuttling COVID-19 cases to certain hospitals to balance system-wide patient burden, may also imbalance case severity across hospitals and care management between hospitals. This was ultimately a major reason to restrict the model training to a single center and perform testing in other hospital centers. Additionally, in this paper, we present outcome classification derived from a learned optimization threshold cutoff. Further work is needed to identify clinically relevant thresholds for classifying predicted probabilities. Finally, although XGBoost is superior to other models in handling missing data, a notable drawback is its bias toward continuous features instead of categorical ones [49]. However, collinearities between some categorical features in this data set may be present with other continuous features, as exhibited by the covariance strength between hypertension and systolic blood pressure and creatinine in Figure S1 in Multimedia Appendix 2, which can then serve as vehicles for capturing these categorical pieces of information.

Conclusions

The COVID-19 pandemic unequivocally represents an unprecedented public health crisis. Health care institutions are facing extreme difficulties in managing resources and personnel. Physicians are treating record numbers of patients and are continuously exposing themselves to a highly contagious and virulent disease with varying symptomatology. Only a few therapeutic options have demonstrated improvement in patient outcomes. Our externally and prospectively validated models successfully predict critical illness and mortality up to 10 days in advance in a diverse patient population from admission information alone. We believe that this model also identified important markers for acute care prognosis that can be used by health care institutions to improve care decisions at both the physician and hospital level for management of COVID-19–positive patients.

Authors' Contributions

BSG, GNN, AWC, and ACJ conceived and designed the study. AV, SS, JKDF, FFC, IP, KWJ, SJL, RM, MD, SZ, FR, NDB, NN, AK, and PT collected the data. PG, EG, MS, DM, PK, JF, RMF, EA, AK, JAA, EB, CRH, BM, EJN, EES, JHC, CCC, VF, DSC, RLR, EPB, MAL, JN, and ZAF facilitated data infrastructure and access. AV, SS, AJK, JKDF, FFC, IP, KWJ, SJL, RM, SZ, FR, NDB, NN, and FR were involved in data analysis. AV, SS, AJK, JKDF, FFC, IP, KWJ, SJL, RM, SZ, FR, NDB, NN, FR, KA, AK, PM, AL, MP, PFO, LHH, BP, and MAL were involved in interpreting the results. AV, SS, AJR, JKDF, FR, NN, ACJ, GNN, and BSG drafted the initial manuscript. ACJ, AWC, GNN and BSG served as co-corresponding authors. All authors provided critical comments and edited, revised, and approved the manuscript in its final form for submission.

Conflicts of Interest

SS is a cofounder and equity owner of Monogram Orthopedics. KWJ received fees from and holds equity in Tempus Labs. JAA received research grants and personal fees from Gilead, Merck, Janssen, and Viiv; personal fees from Medicure and Theratechnologies; and research grants from Atea, Pfizer and Regeneron, all outside of the submitted work. ES is the founding CEO and equity owner of Sema4.

Multimedia Appendix 1
Supplementary Table 1: Final XGBoost, LASSO and logistic regression model hyperparameters.
[DOCX File , 15 KB - jmir_v22i11e24018_app1.docx]

Multimedia Appendix 2
Supplementary figures.
[DOCX File , 1913 KB - jmir_v22i11e24018_app2.docx]

Multimedia Appendix 3
Supplementary Table 2: Brier scores for each model and calibration type.
[DOCX File , 20 KB - jmir_v22i11e24018_app3.docx]

Multimedia Appendix 4
Supplementary Table 3: TRIPOD guidelines.
[DOCX File , 22 KB - jmir_v22i11e24018_app4.docx]

Multimedia Appendix 5
Supplementary Table 4: Baseline feature variability across all patients.
[DOCX File , 35 KB - jmir_v22i11e24018_app5.docx]

Multimedia Appendix 6
Supplementary Table 5: Model performance by experiment.
[DOCX File , 26 KB - jmir_v22i11e24018_app6.docx]

Multimedia Appendix 7
Supplementary Table 6: Clinical features ranked in decreasing order based on their significance (most significant predictor=1, least significant predictor=10) for critical event and mortality at 7 days.
[DOCX File , 14 KB - jmir_v22i11e24018_app7.docx]

References

1. Coronavirus disease (COVID-19) Situation Report – 191. World Health Organization. 2020 Jul 29. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200729-covid-19-sitrep-191.pdf?sfvrsn=2c327e9e_2 [accessed 2020-10-21]

2. Guan W, Ni Z, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]

3. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. JAMA 2020 Apr 14;323(14):1335. [doi: 10.1001/jama.2020.4344] [Medline: 32181795]

4. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]

5. Chan L, Chaudhary K, Saha A, Chauhan K, Vaid A, Zhao S, Mount Sinai COVID Informatics Center (MSCIC), et al. AKI in Hospitalized Patients with COVID-19. J Am Soc Nephrol 2020 Sep 03 [FREE Full text] [doi: 10.1681/ASN.2020050615] [Medline: 32883700]

6. Lala A, Johnson K, Januzzi HL, Russak AJ, Paranjpe I, Richter F, Mount Sinai COVID Informatics Center. Prevalence and Impact of Myocardial Injury in Patients Hospitalized With COVID-19 Infection. J Am Coll Cardiol 2020 Aug 04;76(5):533-546 [FREE Full text] [doi: 10.1016/j.jacc.2020.06.007] [Medline: 32517963]

7. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill 2020 Mar;25(10):pii=2000180 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000180] [Medline: 32183930]

8. Santosh K. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. J Med Syst 2020 Mar 18;44(5):93 [FREE Full text] [doi: 10.1007/s10916-020-01562-1] [Medline: 32189081]

9. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328. [doi: 10.1136/bmj.m1328] [Medline: 32265220]

10. Chen W, Zheng K, Liu S, Yan Z, Xu C, Qiao Z. Plasma CRP level is positively associated with the severity of COVID-19. Ann Clin Microbiol Antimicrob 2020 May 15;19(1):18 [FREE Full text] [doi: 10.1186/s12941-020-00362-2] [Medline: 32414383]

11. Liu F, Li L, Xu M, Wu J, Luo D, Zhu YS, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. J Clin Virol 2020 Jun;127:104370 [FREE Full text] [doi: 10.1016/j.jcv.2020.104370] [Medline: 32344321]

12. Liu J, Liu Y, Xiang P, Pu L, Xiong H, Li C, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J Transl Med 2020 May 20;18(1):206 [FREE Full text] [doi: 10.1186/s12967-020-02374-0] [Medline: 32434518]

13. Petrilli C, Jones S, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020 May 22;369:m1966. [doi: 10.1136/bmj.m1966] [Medline: 32444366]

14. Yadaw A, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. Lancet Digit Health 2020 Oct 22;2(10):e516-e525 [FREE Full text] [doi: 10.1016/S2589-7500(20)30217-X]

15. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv. Preprint posted online on March 17, 2020 . [doi: 10.1101/2020.02.27.20028027]

16. Zhou Y, He Y, Yang H, Yu H, Wang T, Chen Z, et al. Development and validation a nomogram for predicting the risk of severe COVID-19: A multi-center study in Sichuan, China. PLoS One 2020;15(5):e0233328 [FREE Full text] [doi: 10.1371/journal.pone.0233328] [Medline: 32421703]

17. Menni C, Valdes A, Freidin M, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]

18. Chow D, Glabis-Bloom J, Soun J, Weinberg B, Berens-Loveless T, Xie X, et al. Development and External Validation of a Prognostic Tool for COVID-19 Critical Disease. medRxiv. Preprint posted online on May 11, 2020 . [doi: 10.1101/2020.05.06.20093435]

19. Kinar Y, Lanyado A, Shoshan A, Yesharim R, Domany T, Shalev V, et al. Predicting individual risk for COVID19 complications using EMR data. medRxiv. Preprint posted online on June 05, 2020 . [doi: 10.1101/2020.06.03.20121574]

20.   Levy T, Richardson S, Coppa K, Barnaby DP, McGinn T, Becker LB, et al. Development and Validation of a Survival Calculator for Hospitalized Patients with COVID-19. medRxiv. Preprint posted online on June 02, 2020 [FREE Full text] [doi: 10.1101/2020.04.22.20075416] [Medline: 32511640]

21.   Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun 2020 Jul 15;11(1):3543 [FREE Full text] [doi: 10.1038/s41467-020-17280-8] [Medline: 32669540]

22.   Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical Predictive Models for COVID-19: Systematic Study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]

23.   About Race. US Census Bureau. URL: https://www.census.gov/topics/population/race/about.html [accessed 2020-10-22]

24.   Alger H, Williams J, Walchok J, Bolles M, Fonarow G, Rutan C. Role of Data Registries in the Time of COVID-19. Circ Cardiovasc Qual Outcomes 2020 May;13(5):e006766. [doi: 10.1161/CIRCOUTCOMES.120.006766] [Medline: 32298145]

25.   Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]

26.   Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci 2020 Mar;63(3):364-374 [FREE Full text] [doi: 10.1007/s11427-020-1643-8] [Medline: 32048163]

27.   Zhu J, Ji P, Pang J, Zhong Z, Li H, He C, et al. Clinical characteristics of 3062 COVID-19 patients: A meta-analysis. J Med Virol 2020 Apr 15:2020 [FREE Full text] [doi: 10.1002/jmv.25884] [Medline: 32293716]

28.   Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Aug Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]

29.   Lundberg S, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]

30.   Collins G, Reitsma J, Altman D, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015 Jan 06;13(3):1-58 [FREE Full text] [doi: 10.1186/s12916-014-0241-z] [Medline: 25563062]

31.   COVID-ML-Prediction. GitHub. URL: https://github.com/HPIMS/COVID-ML-Prediction [accessed 2020-10-22]

32.   Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017 Presented at: Neural Information Processing Systems 2017; December 4-9, 2017; Long Beach, CA p. 4765-4774 URL: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

33.   Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 May;46(5):846-848 [FREE Full text] [doi: 10.1007/s00134-020-05991-x] [Medline: 32125452]

34.   Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and Laboratory Predictors of In-Hospital Mortality in 305 Patients with COVID-19: A Cohort Study in Wuhan, China. SSRN Journal. Preprint posted online on March 05, 2020 . [doi: 10.2139/ssrn.3546115]

35.   Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020 Apr;8(4):e21. [doi: 10.1016/s2213-2600(20)30116-8]

36.   Kim N, Ha E, Moon J, Lee Y, Choi E. Acute Hyperglycemic Crises with Coronavirus Disease-19: Case Reports. Diabetes Metab J 2020 Apr;44(2):349-353 [FREE Full text] [doi: 10.4093/dmj.2020.0091] [Medline: 32347027]

37.   Orioli L, Hermans M, Thissen J, Maiter D, Vandeleene B, Yombi JC. COVID-19 in diabetic patients: Related risks and specifics of management. Ann Endocrinol (Paris) 2020 Jun;81(2-3):101-109 [FREE Full text] [doi: 10.1016/j.ando.2020.05.001] [Medline: 32413342]

38.   Li H, Liu L, Zhang D, Xu J, Dai H, Tang N, et al. SARS-CoV-2 and viral sepsis: observations and hypotheses. Lancet 2020 May;395(10235):1517-1520. [doi: 10.1016/s0140-6736(20)30920-x]

39.   Drent M, Cobben N, Henderson R, Wouters E, van Dieijen-Visser M. Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung damage or inflammation. Eur Respir J 1996 Aug;9(8):1736-1742. [doi: 10.1183/09031936.96.09081736] [Medline: 8866602]

40.   Henry B, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med 2020 Jun 25;58(7):1021-1028. [doi: 10.1515/cclm-2020-0369] [Medline: 32286245]

41.   Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, et al. Hematological findings and complications of COVID-19. Am J Hematol 2020 Jul;95(7):834-847 [FREE Full text] [doi: 10.1002/ajh.25829] [Medline: 32282949]

42.   Lim J, Park S, Jeon Y, Cho JH, Jung HY, Choi JY, et al. Fatal Outcomes of COVID-19 in Patients with Severe Acute Kidney Injury. J Clin Med 2020 Jun 03;9(6):1718 [FREE Full text] [doi: 10.3390/jcm9061718] [Medline: 32503180]

43. Ronco C, Reis T, Husain-Syed F. Management of acute kidney injury in patients with COVID-19. Lancet Respir Med 2020 Jul;8(7):738-742. [doi: 10.1016/s2213-2600(20)30229-0]

44. Kochi A, Tagliari A, Forleo G, Fassini G, Tondo C. Cardiac and arrhythmic complications in patients with COVID-19. J Cardiovasc Electrophysiol 2020 May;31(5):1003-1008 [FREE Full text] [doi: 10.1111/jce.14479] [Medline: 32270559]

45. Klok F, Kruip M, van der Meer NJM, Arbous MS, Gommers DAMPJ, Kant KM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res 2020 Jul;191:145-147 [FREE Full text] [doi: 10.1016/j.thromres.2020.04.013] [Medline: 32291094]

46. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. N Engl J Med 2020 Apr 23;382(17):e38. [doi: 10.1056/nejmc2007575]

47. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, et al. Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. JAMA Cardiol 2020 Jul 01;5(7):802-810 [FREE Full text] [doi: 10.1001/jamacardio.2020.0950] [Medline: 32211816]

48. Sigel K, Swartz T, Golden E, Paranjpe I, Somani S, Richter F, et al. Covid-19 and People with HIV Infection: Outcomes for Hospitalized Patients in New York City. Clin Infect Dis 2020 Jun 28:ciaa880 [FREE Full text] [doi: 10.1093/cid/ciaa880] [Medline: 32594164]

49. Strobl C, Boulesteix A, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007 Jan 25;8:25 [FREE Full text] [doi: 10.1186/1471-2105-8-25] [Medline: 17254353]

Abbreviations

ABG: arterial blood gas
AKI: acute kidney injury
AUC-ROC: area under the receiving operator characteristic curve
AU-PRC: area under the precision-recall curve
BMP: basic metabolic panel
EHR: electronic health record
ICU: intensive care unit
ICD-9/10-CM: International Classification of Diseases-9/10-Clinical Modification
kNN: k-nearest neighbors
LASSO: least absolute shrinkage and selection operator
LDH: lactate dehydrogenase
LR: logistic regression
MSB: Mount Sinai Brooklyn
MSH: Mount Sinai Hospital
MSHS: Mount Sinai Health System
MSM: Mount Sinai Morningside
MSQ: Mount Sinai Queens
MSW: Mount Sinai West
OH: other hospitals
RDW: red cell distribution width
RT-PCR: reverse transcriptase–polymerase chain reaction
SHAP: SHapley Additive exPlanation
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
VBG: venous blood gas
XGBoost: Extreme Gradient Boosting

XSL•FO
RenderX

XSL•FO

RenderX

<u>Original Paper</u>

Concerns About Information Regarding COVID-19 on the Internet: Cross-Sectional Study

Yusui Zhao[1], MSc; Shuiyang Xu[1], BA; Lei Wang[1], MSc; Yu Huang[1], MSc; Yue Xu[1], MSc; Yan Xu[1], MSc; Qiaohong Lv[1], MSc; Zhen Wang[1*], MSc; Qingqing Wu[1*], MSc

Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
[*]these authors contributed equally

Corresponding Author:
Qingqing Wu, MSc
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road
Hangzhou, 310051
China
Phone: 86 571 87115239
Email: qqwu@cdc.zj.cn

Abstract

Background: Since the outbreak of COVID-19, the Chinese government and the Chinese Center for Disease Control and Prevention have released COVID-19–related information to the public through various channels to raise their concern level of the pandemic, increase their knowledge of disease prevention, and ensure the uptake of proper preventive practices.

Objective: Our objectives were to determine Chinese netizens' concerns related to COVID-19 and the relationship between their concerns and information on the internet. We also aimed to elucidate the association between individuals' levels of concern, knowledge, and behaviors related to COVID-19.

Methods: The questionnaire, which consisted of 15 closed-ended questions, was designed to investigate Chinese netizens' knowledge about COVID-19. The self-selection online survey method of nonprobability sampling was used to recruit participants through Dingxiangyisheng WeChat (a public, medical, and health service platform in China) accounts. Standard descriptive statistics and multivariate logistic regression analyses were conducted to analyze the data.

Results: In total, 10,304 respondents were surveyed on the internet (response rate=1.75%; 10,304/590,000). Nearly all (n=9803, 95.30%) participants were concerned about "confirmed cases" of COVID-19, and 87.70% (n=9036) received information about the outbreak through social media websites. There were significant differences in participants' concerns by sex (P=.02), age (P<.001), educational attainment (P=.001), and occupation (P<.001). All knowledge questions and preventive practices were associated with concerns about COVID-19. The results of the multivariate logistic regression indicated that participants' sex, educational attainment, occupation and employment status, knowledge acquisition, and concern level were significantly associated with the practice of proper preventive behaviors.

Conclusions: This study elucidated Chinese netizens' concerns, information sources, and preventive behaviors related to the COVID-19 pandemic. Sex, educational attainment, occupation and employment status, knowledge acquisition, and level of concern were key factors associated with proper preventive behaviors. This offers a theoretical basis for the government to provide targeted disease prevention and control information to the public.

(*J Med Internet Res 2020;22(11):e20487*)   doi:10.2196/20487

KEYWORDS

coronavirus; COVID-19; disease prevention; internet; knowledge

Introduction

In December 2019, the first cases of COVID-19 were detected in Wuhan, Hubei Province, China. Soon after, the pandemic spread rapidly across China and abroad [1]. As a result of the rapidly increasing numbers of confirmed cases and deaths, the public paid unprecedented attention to the pandemic [2]. By February 1, 2020, 14,380 cases and 304 deaths had been confirmed in China [3].

Internet access has increased worldwide during the past decade, reaching 48% of the world population in 2017 [4]. As of December 31, 2018, there were 829 million internet users in China (a penetration rate of 59.6%) [5]. Compared with severe acute respiratory syndrome (SARS) in 2003 [6], the Middle Eastern respiratory syndrome (MERS) [7], and the Zika virus [8], people are more willing to learn about the current pandemic and acquire relevant protection knowledge over the internet [9]. As there is currently neither a vaccine nor a specific drug treatment for COVID-19, performing personal protection behaviors is still the most effective measure to control the pandemic.

Local governments have responded with a series of interventions to raise public concerns [10], including raising the level of emergency response [11] and traffic restrictions [12]. Concurrently, disease prevention and control organizations have employed varied means of health education to encourage the public's preventive behaviors. For example, the Chinese Center for Disease Control and Prevention (China CDC) issued COVID-19 Guidelines for Public Protection (version 2). A printed edition was distributed in the community, and an downloadable electronic version was made available on the official website [13]. Further, the local mainstream media, websites, and a 24-hour telephone hotline disseminated information about behaviors to prevent respiratory communicable diseases. Previous research has found that the more the public pays attention to public health emergencies, the easier it will be for them to take the initiative to seek knowledge regarding disease-related prevention and control, adopt correct personal protective behaviors, and improve support for local disease-related prevention and control policies [14,15]. There is also a large number of scientific literature that reports on people's knowledge, attitudes, and behaviors related to infectious diseases [6,8]. However, to the best of our knowledge, there is limited research addressing the relationships between the public's level of concern and knowledge of and behaviors related to infectious disease prevention [16]. Based on the above theory, our hypothesis is that there may be a certain association between the public's concerns, knowledge, and preventive practices related to COVID-19. In addition, raising the level of public concern regarding the risks related to COVID-19 may be an important prerequisite in influencing the public's understanding, and may be a key entry point in enhancing self-protection.

Hence, our main objectives were to determine Chinese netizens' concerns related to COVID-19 and its relationship with internet content, and to elucidate the association between individuals' concerns, knowledge, and preventive practices related to COVID-19.

Methods

Design

Following Facebook and Twitter's success in medical education in Western universities [17], WeChat is a new media app that is currently recognized as the best social networking site to use for a similar purpose, given that its users span all ages and professions in China [18]. There are currently an estimated 963 million WeChat users in China [5]. As a result, WeChat has been widely utilized in medical fields and has had a positive influence. Dingxiangyisheng WeChat is a public, medical, and health service platform in China. It mainly provides reliable online consultation services, health knowledge lectures, professional science articles, and other services for the domestic public. As the largest professional medical website in China, it has more than 5.5 million users and receives millions of daily visitors. Its users are mainly those who care about their health and that of their families, especially young Chinese mothers.

This study used opportunity sampling. A message with a questionnaire related to COVID-19 was sent to all users with a Dingxiangyisheng WeChat account. Individuals who were users of a WeChat account could participate in this study through their mobile phone or computer. In summary, 590,000 users read this information, and 10,304 of them volunteered to participate in this survey. To avoid repeated questionnaire submission, each respondent could only submit the questionnaire once per registered phone number. Questionnaires could not be revised or repeated after submission. This study was approved by the Ethics Committee of the Zhejiang Provincial CDC.

Participants

A cross-sectional, population-based internet survey was conducted from January 31 to February 2, 2020, at the beginning of the COVID-19 pandemic in China [19]. Participants were aged ≥15 years and lived in China. Participation was voluntary without any financial incentive. Participants' personal information was kept confidential and stored at Dingxiangyisheng. Participants who did not fully complete the questionnaire were excluded.

Questionnaire

The online questionnaire, designed by health educational experts from the Zhejiang CDC, consisted of 15 closed-ended questions that aimed to collect the following information from respondents (Multimedia Appendix 1):

1. Sociodemographic and background information, including age, sex, province, occupation and employment status, educational attainment, and infection status (6 questions).
2. Concerns related to COVID-19 (1 question). A 3-point Likert-type scale was used to ascertain the level of concern (1=very, 2=general, and 3=not at all).
3. Participants' actual knowledge of COVID-19, including four aspects: its incubation period, transmission routes, symptoms, and knowledge of personal disease prevention behaviors (4 questions).
4. Practices of preventive measures against COVID-19, including mask wearing, handwashing with soap at home, avoiding public places and public transportation, and proper way to sneeze (4 questions).

Prior to administering the online survey, the questionnaire was pilot tested using the interception sampling method. Inspectors stopped passersby at a regular intersection and asked if they would be willing to answer the survey, which took approximately 10 minutes to complete. The survey questionnaire was pilot tested in January 2020 (N=100) to ensure practicability and interpretability of answers. For our online survey

XSL•FO
RenderX

(N=10,304), the overall Cronbach α value was 0.771. The Kaiser-Meyer-Olkin test coefficient was 0.578. By factor analysis, the cumulative contribution rate of the four factors with a characteristic root greater than 1 was 44.866, indicating that the questionnaire has good reliability and validity.

Data Analyses

Data were analyzed using SPSS, version 13.0 (IBM Corp). Sociodemographic and background information was converted into classification variables or grade variables, and the corresponding composition ratio was calculated. For questions about knowledge and practices of preventive measures, answers of "yes" were coded into the correct knowledge and practices, while answers of "no" and "unclear" were coded into incorrect knowledge and practices. The nonparametric Mann-Whitney test was used for comparisons among groups with different characteristics, and the Kruskal-Wallis H test was used for comparisons among multiple groups. Questions regarding knowledge and personal protective measures were expressed by dichotomous variables. Mann-Whitney U tests of two independent samples were conducted to identify apparent associations between individuals' concerns with knowledge and practices of preventive measures. A P value <.05 was considered significant.

Results

Participants' Characteristics

A total of 10,304 individuals completed the online survey (response rate=1.75%; 10,304/590,000). Participants' sociodemographic characteristics are shown in Table 1.

Table 1. Participants' sociodemographic characteristics (N=10,304).

Variable	Participants, n (%)
Sex	
Male	2670 (25.91)
Female	7634 (74.09)
Age (years)	
15-20	901 (8.74)
21-30	4830 (46.88)
31-40	2945 (28.58)
41-50	1141 (11.07)
51-60	403 (3.91)
≥60	84 (0.82)
Educational attainment	
No education	92 (0.89)
Primary	434 (4.21)
Preparatory	1117 (10.84)
Secondary	7219 (70.06)
University graduate	1442 (13.99)
Occupation and employment status	
Government and public institution staff	1729 (16.78)
Enterprise/commercial/service worker	4100 (39.79)
Farmer	104 (1.01)
Retired	202 (1.96)
Homemaker	751 (7.29)
Student	1894 (18.38)
Unemployed	395 (3.83)
Medical staff	668 (6.48)
Other	461 (4.47)
Infection status	
Confirmed case	0 (0)
Suspected case	29 (0.28)
Close contact with a confirmed case	293 (2.84)
None of the above	9982 (96.88)

Participants' Concerns and Their Information Sources

Table 2 presents data regarding participants' information sources and what they were concerned about. Nearly all participants were concerned about the number of confirmed cases. Most of the information about the outbreak was received through a social media website, while far less was obtained through traditional mass media (eg, television and newspapers).

Table 2. Participants' concerns and information sources related to COVID-19.

Variable	Participants, n (%)
Concern	
Confirmed cases	9803 (95.30)
Number of deaths	8463 (82.27)
Number of recovered	8131 (79.04)
Suspected cases	6942 (67.48)
Number of severe patients	4795 (46.61)
Information source	
Social media website	9036 (87.70)
Smartphone app	4556 (44.22)
Television	4054 (39.35)
Friends/relatives	2447 (23.75)
Government website	2201 (21.36)
Short message service	1111 (10.78)
Community outreach	1070 (10.39)
Newspaper	679 (6.59)
Significant other/partner	166 (1.61)

Concerns According to Participants' Characteristics

Since the number of participants who reported they were "unconcerned" was very small, we merged the "unconcerned" and "general" categories, resulting in two categories for level of concern: "very" concerned or "general/unconcerned." Table 3 presents the concerns according to participants' characteristics. There were significant differences in concerns by sex, age, educational attainment, and occupation—specifically, women, those who were older, those who were more educated, and those who were retired expressed a greater level of concern than their counterparts. Students were least concerned compared to other occupation types. There were marginally significant differences between infection status and concerns; those with close contact with a confirmed case were more concerned about COVID-19 than the other types listed in Table 3.

XSL•FO

RenderX

Table 3. Concerns about COVID-19 according to participants' characteristics.

Variable	Very (n=8112), n (%)	General/not at all (n=2177), n (%)	Z or χ^2	P value
Sex			χ^2=2.42	.02
Male	2146 (80.37)	524 (19.63)		
Female	5966 (78.15)	1668 (21.85)		
Age (years)			Z=318.28	<.001
15-20	610 (67.70)	291 (32.30)		
21-30	3542 (73.33)	1288 (26.67)		
31-40	2517 (85.47)	428 (14.53)		
41-50	1004 (87.99)	137 (12.01)		
51-60	364 (90.32)	39 (9.68)		
≥60	75 (89.29)	9 (9.71)		
Educational attainment			Z=11.97	.001
No education	70 (76.09)	20 (21.91)		
Primary	340 (78.34)	93 (21.66)		
Preparatory	846 (75.74)	270 (24.26)		
Secondary	5678 (78.65)	1530 (21.34)		
University graduate	1178 (81.69)	264 (18.31)		
Occupation and employment status			Z=35.50	<.001
Government and public institution staff	1461 (84.50)	268 (15.50)		
Enterprise/commercial/service worker	3234 (78.88)	866 (21.12)		
Farmer	85 (81.73)	19 (18.27)		
Retired	181 (89.60)	21 (10.40)		
Homemaker	613 (81.62)	138 (18.38)		
Student	1300 (68.64)	594 (31.36)		
Unemployed	286 (72.41)	109 (27.59)		
Medical staff	572 (85.63)	96 (14.37)		
Other	380 (82.43)	81 (17.57)		
Infection status			Z=5.83	.054
Suspected case	18 (62.07)	11 (37.93)		
Close contact	234 (79.86)	59 (20.14)		
None of the above	7860 (78.74)	2122 (21.26)		

Concerns About COVID-19 According to Related Knowledge and Preventive Practices

For all knowledge questions, participants who were more (vs less) concerned about COVID-19 were significantly more likely to master relevant knowledge. For almost all practice-related questions, participants who had more (vs less) concerns about COVID-19 were significantly more likely to take the right action to avoid SARS-CoV-2 infection. Notably, less concerned participants were more likely to avoid public places and public transportation than other participants (Table 4).

Table 4. Concerns about COVID-19 according to related knowledge and preventive practices.

Variable	Very, n (%)	General/not at all, n (%)	Total, n (%)	χ^2	P value
Knowledge items (proportion answered correctly)					
Incubation period	7041 (86.80)	1829 (83.51)	8870 (86.08)	16.24	<.001
Transmission routes	6212 (76.58)	1619 (74.05)	7831 (76.00)	6.99	.009
Symptoms	6675 (82.29)	1663 (76.21)	8338 (80.92)	46.05	<.001
Personal preventive knowledge	7020 (86.54)	1826 (83.37)	8846 (85.85)	14.87	<.001
Preventive practices					
Knows how to wear a mask when going out	7916 (97.58)	2111 (96.42)	10,027 (97.31)	10.79	<.001
Washes hands after getting home	7548 (93.05)	1878 (85.90)	9426 (91.48)	120.32	<.001
Avoids going to public places and using public transportation	2160 (26.63)	519 (23.61)	2679 (26.00)	7.81	.005
Can cough or sneeze properly	4822 (59.44)	1142 (52.18)	5964 (57.88)	38.18	<.001

Multivariate Logistic Regression

For the analysis, respondents who answered the 4 knowledge questions correctly were deemed to have mastered COVID-19 preventive knowledge; otherwise, they were deemed to have not mastered it. Similarly, respondents who correctly answered the 4 questions related to preventive practices were regarded as having correct preventive practices, while those who did not were regarded as not having correct preventive practices. Preventive practices were used as the dependent variable, while sex, age, educational attainment, occupational status, and knowledge acquisition were used as independent variables in the multivariate regression analysis. The multivariate logistic regression revealed the factors that contributed to having correct preventive practices (Table 5). Women were more likely than men (P=.001) to practice the correct precautions. Participants with more education (secondary, P=.03; university graduate, P=.01) were more likely to perform the right preventive practices than those with less educational attainment. Students were less likely to take the right preventive practices compared to other occupational groups (P=.02). Concern (P<.001) and knowledge acquisition (P=.001) regarding the COVID-19 outbreak were significantly associated with proper preventive practices.

Table 5. Multivariate logistic regression of factors that contributed to performing preventive practices.

Variables	B	SE	Wald χ^2	P value	Odds ratio	95% CI
Sex (reference: male)						
Female	–0.22	0.07	10.84	.001	0.80	0.70-0.91
Educational attainment (reference: no education)						
Secondary	–0.26	0.12	4.71	.03	0.77	0.61-0.98
University graduate	–0.20	0.08	6.01	.01	0.82	0.70-0.96
Occupation and employment status (reference: government and public institution staff)						
Student	0.38	0.17	5.19	.02	1.46	1.05-2.02
Knowledge acquisition (reference: no)						
Yes	–0.19	0.06	10.98	.001	0.83	0.74-0.93
Concern (reference: very)						
General/not at all	0.48	0.08	38.16	<.001	1.61	1.39-1.88

Discussion

Principal Findings

This study elucidated the concerns and information sources related to the COVID-19 pandemic of Chinese netizens. We also determined the relationship between these individuals' concerns and their personal protection knowledge and behaviors.

Our results revealed that the vast majority of the sample were very concerned about this outbreak. Among all surveyed participants, 78.73% were "very concerned," which is significantly higher than that found in a study involving Chinese residents during the SARS pandemic [20]. On the one hand, as the pandemic developed rapidly in early February, the Chinese government and relevant departments took a variety of strict prevention and control measures, including disseminating information through various media platforms [10,12]. On the other hand, with the continuous popularization of the internet in China, the public has more opportunities to obtain information

XSL•FO
RenderX

that directly concerns them than ever before, which motivates them to actively take preventive measures [21].

The number of confirmed COVID-19 cases per day was the item that nearly all participants were concerned about. In response, China's National Health Commission and health departments at all levels announce the number of confirmed cases in cities and counties daily through official media websites to keep the public informed of new outbreaks in the region. With the rapid development of the internet worldwide, the way the public obtains health information has changed greatly since the SARS [6], H7N9 [22], and Zika [23] outbreaks. Specifically, the proportion of traditional media, such as television and newspapers, has gradually decreased, while the proportion of new media, such as websites and smartphone apps, has gradually increased [24]. This suggests that, when there is a new infectious disease outbreak, the government should release information to the public through an official website and other relevant media to increase public awareness as soon as possible, and ensure that the correct information is disseminated to foster protective knowledge and behavior.

Consistent with previous studies [25], we found that most people have the right knowledge to prevent COVID-19, especially those who were "very" concerned about the pandemic. Surprisingly, we found that the percentage of respondents electing to avoid going to public places and using public transportation was "very low" compared to other personal preventive practices. It is worth noting that the outbreak happened during the Spring Festival holiday in China; thus, population mobility had increased. Even in the face of such a severe outbreak, most people did not avoid going to public places or using public transportation [12].

As in previous studies [24,26], we found a statistically significant association between sex and personal preventive practices. Women are more likely than men to take the right preventive practices. Although a univariate analysis found significant statistical differences between age and personal preventive practices, multivariate logistic regression analysis did not find such differences. This disparity in the results might be explained by the collinearity between age and educational attainment when we conducted the multivariate logistic regression analysis. Students were significantly less concerned about the outbreak than other occupational groups. Moreover, students were less likely to take proper preventive practices than other occupational groups. Students who were less interested in acquiring information about diseases and utilizing protective measures would increase their risk of infection [8]. Thus, before returning to school, students should be educated about the transmission of COVID-19 and how to develop good personal hygiene habits.

Increased attention to the pandemic was associated with increased knowledge about prevention and control, which coincides with previous results [27]. This shows that knowledge acquisition and the adoption of appropriate behaviors are closely related to individuals' disease-related concerns. For individuals who are not concerned, stakeholders need to increase health education and ensure they receive the appropriate information.

Strengths and Limitations

The strength of our study is its large sample, which was recruited during a critical period—the early stages of the COVID-19 outbreak. Similar to the results of a recent online survey about COVID-19 [28], our sample was overrepresentative of women, well-educated people, and white-collar workers.

There are several limitations to this study. First, owing to limited access to the internet and online health information resources, certain populations such as the elderly and rural residents were insufficiently represented; therefore, this study is not representative of all Chinese people. The study, which recruited participants through Dingxiangyisheng WeChat, had a bias in favor of women, especially young Chinese mothers. Second, we employed a cross-sectional design; therefore, we only revealed correlations between variables and cannot infer causal links between them. More in-depth scientific research is needed.

Conclusions

This study elucidated the concerns, information sources, and preventive behaviors of Chinese netizens related to the COVID-19 pandemic. Sex, educational attainment, occupation and employment status, knowledge acquisition, and concern level were key factors associated with proper preventive behaviors. This offers a theoretical basis for the government to provide targeted disease prevention and control information disseminated to the public.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COVID-19 prevention and treatment questionnaire in Chinese.
[DOC File , 355 KB - jmir_v22i11e20487_app1.doc]

References

XSL•FO
RenderX

1.   Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]

2.   Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. The Lancet 2020 Mar;395(10226):764-766. [doi: 10.1016/s0140-6736(20)30421-9]

3.   Update on pneumonia outbreak by novel coronavirus as of 24 o'clock on February 1, 2020. National Health Commission, People's Republic of China. 2020 Feb 2. URL: http://www.nhc.gov.cn/xcs/yqtb/202002/d5c495da742f4739b7f99339c3bd032f.shtml [accessed 2020-02-02]

4.   The state of broadband 2017: broadband catalyzing sustainable development. International Telecommunications Union/UNESCO. 2017 Sep. URL: https://www.itu.int/dms_pub/itu-s/opb/pol/s-pol-broadband.18-2017-pdf-e.pdf [accessed 2020-03-26]

5.   The 43rd statistical report on Internet development in China. China Internet Network Information Center. 2019 Feb. URL: http://www.cac.gov.cn/wxb_pdf/0228043.pdf [accessed 2019-07-17]

6.   Vartti AM, Oenema A, Schreck M, Uutela A, de Zwart O, Brug J, et al. SARS knowledge, perceptions, and behaviors: a comparison between Finns and the Dutch during the SARS outbreak in 2003. Int J Behav Med 2009;16(1):41-48 [FREE Full text] [doi: 10.1007/s12529-008-9004-6] [Medline: 19184625]

7.   Alsahafi A, Cheng A. Knowledge, Attitudes and Behaviours of Healthcare Workers in the Kingdom of Saudi Arabia to MERS Coronavirus and Other Emerging Infectious Diseases. Int J Environ Res Public Health 2016 Dec 06;13(12):1214 [FREE Full text] [doi: 10.3390/ijerph13121214] [Medline: 27929452]

8.   Plaster AN, Painter JE, Tjersland DH, Jacobsen KH. University Students' Knowledge, Attitudes, and Sources of Information About Zika Virus. J Community Health 2018 Aug 9;43(4):647-655. [doi: 10.1007/s10900-017-0463-z] [Medline: 29318503]

9.   Fan KS, Ghani SA, Machairas N, Lenti L, Fan KH, Richardson D, et al. COVID-19 prevention and treatment information on the internet: a systematic analysis and quality assessment. BMJ Open 2020 Sep 10;10(9):e040487 [FREE Full text] [doi: 10.1136/bmjopen-2020-040487] [Medline: 32912996]

10.   Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642. [doi: 10.1126/science.abb6105] [Medline: 32234804]

11.   Tibet activates highest-level public health alert. Chinadaily. 2020 Jan 29. URL: http://www.chinadaily.com.cn/a/202001/29/WS5e318a36a3101282172739c1.html [accessed 2020-01-30]

12.   Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]

13.   COVID-19 Guidelines for Public Protection (version 2). Chinese Center For Disease Prevention. 2020. URL: http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11815/202003/P020200306330231617651.pdf [accessed 2020-03-06]

14.   Suri G, Gross JJ. The role of attention in motivated behavior. J Exp Psychol Gen 2015 Aug;144(4):864-872. [doi: 10.1037/xge0000088] [Medline: 26097978]

15.   Tamers SL, Allen J, Yang M, Stoddard A, Harley A, Sorensen G. Does concern motivate behavior change? Exploring the relationship between physical activity and body mass index among low-income housing residents. Health Educ Behav 2014 Dec 30;41(6):642-650 [FREE Full text] [doi: 10.1177/1090198114532289] [Medline: 24786794]

16.   Kennedy CJ, Probart CK, Dorman SM. The relationship between radon knowledge, concern and behavior, and health values, health locus of control and preventive health behaviors. Health Educ Q 1991 Sep 04;18(3):319-329. [doi: 10.1177/109019819101800305] [Medline: 1917508]

17.   Forgie SE, Duff JP, Ross S. Twelve tips for using Twitter as a learning tool in medical education. Medical Teacher 2012 Dec 21;35(1):8-14. [doi: 10.3109/0142159x.2012.746448]

18.   Fujimori M, Morais TC, França E, Toledo ORD, Honório-França A. The attitudes of primary school children to breastfeeding and the effect of health education lectures. J Pediatr (Rio J) 2008 May 30;84(3):224-231. [doi: 10.2223/jped.1791]

19.   Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]

20.   Mingyi Q, Dongmei Y, Wei D, Zheng H, Lili Z, Xinghua L. Behaviour, Cognition and Emotion of the Public in Beijing towards SARS. Chin Ment Health J 2003;17:515-520 [FREE Full text]

21.   Kim Y, Zhong W, Jehn M, Walsh L. Public risk perceptions and preventive behaviors during the 2009 H1N1 influenza pandemic. Disaster Med Public Health Prep 2015 Apr 17;9(2):145-154. [doi: 10.1017/dmp.2014.87] [Medline: 25882121]

22.   Ma X, Liao Q, Yuan J, Liu Y, Liu Y, Chen J, et al. Knowledge, attitudes and practices relating to influenza A(H7N9) risk among live poultry traders in Guangzhou City, China. BMC Infect Dis 2014 Oct 18;14(1):554 [FREE Full text] [doi: 10.1186/s12879-014-0554-8] [Medline: 25324011]

23.   Casapulla SL, Aidoo-Frimpong G, Basta T, Grijalva M. Zika virus knowledge and attitudes in Ecuador. AIMS Public Health 2018;5(1):49-63 [FREE Full text] [doi: 10.3934/publichealth.2018.1.49] [Medline: 30083569]

24.   Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]

25.   Clements JM. Knowledge and Behaviors Toward COVID-19 Among US Residents During the Early Days of the Pandemic: Cross-Sectional Online Questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]

26.   Li S, Feng B, Liao W, Pan W. Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]

27.   Cui B, Liu ZP. Factors Associated with Knowledge and Personal Protective Behavior Related to Avian Influenza A (H7N9) Risk among Chinese Poultry Farmers. Popul Health Manag 2016 Aug;19(4):298-299. [doi: 10.1089/pop.2015.0182] [Medline: 27007253]

28.   Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]

Abbreviations

China CDC: Chinese Center for Disease Control and Prevention
MERS: Middle Eastern respiratory syndrome
SARS: severe acute respiratory syndrome

Original Paper

An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study

Hyung-Jun Kim[1*], MD; Deokjae Han[1*], MD; Jeong-Han Kim[2], MD; Daehyun Kim[3], DDS; Beomman Ha[4], MD, PhD; Woong Seog[4], MD; Yeon-Kyeng Lee[5], PhD; Dosang Lim[5], MSc; Sung Ok Hong[5], PhD; Mi-Jin Park[5], MSc; JoonNyung Heo[4], MD

[1]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Armed Forces Capital Hospital, Seongnam, Republic of Korea

[2]Division of Infectious Diseases, Department of Internal Medicine, Armed Forces Capital Hospital, Seongnam, Republic of Korea

[3]Department of Periodontology, Armed Forces Capital Hospital, Seongnam, Republic of Korea

[4]The Armed Forces Medical Command, Seongnam, Republic of Korea

[5]Division of Chronic Disease Control, Korea Center for Disease Control and Prevention, Cheongju, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
JoonNyung Heo, MD
The Armed Forces Medical Command
81 Saemaeul-ro 177beon-gil, Bundang-gu
Seongnam,
Republic of Korea
Phone: 82 31 725 5490
Email: jnheo@jnheo.com

Abstract

Background: Prioritizing patients in need of intensive care is necessary to reduce the mortality rate during the COVID-19 pandemic. Although several scoring methods have been introduced, many require laboratory or radiographic findings that are not always easily available.

Objective: The purpose of this study was to develop a machine learning model that predicts the need for intensive care for patients with COVID-19 using easily obtainable characteristics—baseline demographics, comorbidities, and symptoms.

Methods: A retrospective study was performed using a nationwide cohort in South Korea. Patients admitted to 100 hospitals from January 25, 2020, to June 3, 2020, were included. Patient information was collected retrospectively by the attending physicians in each hospital and uploaded to an online case report form. Variables that could be easily provided were extracted. The variables were age, sex, smoking history, body temperature, comorbidities, activities of daily living, and symptoms. The primary outcome was the need for intensive care, defined as admission to the intensive care unit, use of extracorporeal life support, mechanical ventilation, vasopressors, or death within 30 days of hospitalization. Patients admitted until March 20, 2020, were included in the derivation group to develop prediction models using an automated machine learning technique. The models were externally validated in patients admitted after March 21, 2020. The machine learning model with the best discrimination performance was selected and compared against the CURB-65 (confusion, urea, respiratory rate, blood pressure, and 65 years of age or older) score using the area under the receiver operating characteristic curve (AUC).

Results: A total of 4787 patients were included in the analysis, of which 3294 were assigned to the derivation group and 1493 to the validation group. Among the 4787 patients, 460 (9.6%) patients needed intensive care. Of the 55 machine learning models developed, the XGBoost model revealed the highest discrimination performance. The AUC of the XGBoost model was 0.897 (95% CI 0.877-0.917) for the derivation group and 0.885 (95% CI 0.855-0.915) for the validation group. Both the AUCs were superior to those of CURB-65, which were 0.836 (95% CI 0.825-0.847) and 0.843 (95% CI 0.829-0.857), respectively.

Conclusions: We developed a machine learning model comprising simple patient-provided characteristics, which can efficiently predict the need for intensive care among patients with COVID-19.

KEYWORDS

COVID-19; machine learning; prognosis; SARS-CoV-2; severe acute respiratory syndrome coronavirus 2

Introduction

COVID-19 is an ongoing global pandemic caused by the severe acute respiratory syndrome virus 2 with over 26 million confirmed cases worldwide as of August 31, 2020 [1]. The virus is highly transmissible [2] and commonly causes symptoms of fever, cough, fatigue, and myalgia [3]. The mortality rate varies from 0.4 to 304.9 deaths per 1000 COVID-19 cases in the United States according to age group [4], while underlying comorbidities and sex are frequently reported as risk factors for a grave prognosis [5,6].

Other than patient factors, the availability of medical resources is also a major factor for higher risk of death by COVID-19 [7]. The reported case fatality rates are higher in areas with sudden upsurges of COVID-19 compared to other regions, even in the same country. In China, the mortality rates were higher in Hubei Province, in which the outbreak sparked, compared to other provinces [8]. In South Korea, the estimated risk of death was 20.8% to 25.9% in Daegu and Gyeongsangbuk-do, which were regions that experienced a sudden COVID-19 outbreak, while other areas had a risk of 1.7% [9]. Such findings are due to the availability of hospital beds, medical professionals, and other necessary supplies. Therefore, prioritizing patients in need of intensive care is crucial to prevent unnecessary consumption of medical resources by mild or asymptomatic patients.

There have been previous efforts to elucidate the risk factors of grave prognoses among patients with COVID-19 [10-13]. A previous report from China used patient demographics, symptoms, comorbidities, lactate dehydrogenase level, neutrophil–lymphocyte ratio, and radiographic abnormality to predict intensive care unit (ICU) admission, invasive ventilation, or death [10]. Another study from Italy concluded that the proportion of well-aerated lungs was associated with ICU admission or death [13]. Other studies from China also emphasized the use of laboratory findings to predict severe types of COVID-19 [11,12]. Although the performance of these models was excellent, they included laboratory or radiographic findings that may not be quickly available in underdeveloped areas. In addition, rapid adjustment of the scoring systems is not feasible when additional data are collected.

In this study, we aimed to develop a prediction model with information that can easily be provided by patients, limited to baseline demographics, comorbidities, and subjective symptoms. The model aimed to predict the need for intensive care among patients with COVID-19 using an automated machine learning (AutoML) technique [14], which can easily adjust the relative importance of different features as further data become available.

Methods

Data Source and Study Population

This was a retrospective study using a nationwide cohort that included all hospitalized patients with COVID-19 in South Korea, developed and managed by the Korean Centers for Disease Control and Prevention. Patients with laboratory-confirmed COVID-19 were either admitted to a hospital or a community treatment center. The Korean Centers for Disease Control and Prevention requested that all hospitals with patients with COVID-19 register and record their patients' data to the cohort. Data were collected retrospectively through medical chart review by the attending physicians in each center, and were uploaded to an online case report form [15].

Among the patients with COVID-19 hospitalized from January 25, 2020, those who died or were released from quarantine as of June 3, 2020, were included in this study. Patients who were admitted until March 20 were assigned to the derivation group, and those hospitalized after March 21 were assigned to the temporal external validation group. The cut-off point of March 20 was arbitrary. However, two major changes occurred in clinical practice during the study period. First, as testing capacity increased during the pandemic, testing criteria were broadened after February 20. Second, services of community treatment centers commenced on March 2, which we used to quarantine patients with mild symptoms. We excluded patients aged <18 years and those with missing data. This study was approved by the Institutional Review Board of the Armed Forces Medical Command (approval number: AFMC-20053-IRB-20-053) with a waiver of informed consent due to the retrospective nature of the study.

Variable Selection

Variables used for developing the machine learning model included information that could easily be provided by patients without the need for laboratory or radiographic evaluation. The variables were age, sex, smoking history, body temperature, underlying comorbidities, activities of daily living (ADL), and symptoms reported by the patients. Comorbidities included diabetes, heart failure, hypertension, asthma, chronic obstructive pulmonary disease, chronic kidney disease, cancer, chronic liver disease, chronic neurological disorders, chronic hematologic disorders, HIV infection, autoimmune diseases, dementia, and pregnancy. The ADL scale was divided into three categories: independent, partially dependent, and totally dependent. Symptoms considered in the cohort were mental status, cough, sputum, hemoptysis, sore throat, rhinorrhea, chest discomfort, myalgia, arthralgia, fatigue, dyspnea, anosmia, headache, vomiting, and diarrhea.

The CURB-65 score, which stands for confusion, urea, respiratory rate, blood pressure, and 65 years of age or older, was chosen as a comparison against the machine learning model [16]. The score consists of mentality, blood urea nitrogen level, respiratory rate, blood pressure, and age [16]. These data were also extracted from the cohort. Levels of blood urea nitrogen were extracted only to calculate the CURB-65 score and were not included in the machine learning model.

Outcome for the Prediction Models

The primary outcome was predicting need for intensive care, which we defined as admission to the ICU, use of extracorporeal

life support, mechanical ventilation, vasopressors, or death during the first 30 days of admission. Information on the use of extracorporeal life support, mechanical ventilation, or vasopressors was included to account for patients who could not be admitted to the ICU due to limited availability.

Machine Learning Analysis

Complete case analysis was performed, and continuous variables were inspected for input errors. AutoML was used to automate the process of constructing pipelines for the development of the machine learning models, such as hyperparameter optimization and model training. H2O.ai was used to develop these AutoML models [14,17].

The algorithms used during the development of the prediction models using AutoML can be classified into three categories: linear, decision tree based, and neural network based. Linear algorithms are essentially multidimensional linear mathematical formulas. They are intuitive and easy to interpret, and problems that can be described in a linear manner would be best solved by these algorithms. Decision tree–based algorithms consist of a multitude of decision trees comprising multiple true or false conditions for input variables. We used the sum of the decisions made by the decision trees for final classification. These models are better for processing categorical variables with multiple levels, and they can account for interactions between variables. A neural network comprises layers of interconnected artificial neurons that are designed based on a biological neuron. These artificial neurons receive multiple inputs that are multiplied by weights, and they output the sum of these inputs. Neural network models are difficult to interpret, but they can successfully represent complicated interactions between inputs. However, these models are not ideal for representing categorical inputs with multiple levels. Since it is unclear which algorithm can best explain the current problem, all these algorithms were used to develop predictive models, which were then compared based on their discriminative power.

The following models were trained in the AutoML process: 3 prespecified XGBoost gradient boosting machine models, a fixed grid of generalized linear models, a default random forest, 5 prespecified H2O gradient boosting machines, a near-default deep neural network, an extremely randomized forest, a random grid of XGBoost gradient boosting machines, a random grid of H2O gradient boosting machines, and a random grid of deep neural network models. Two stacked ensemble models were developed using the aforementioned developed models [18].

Other Statistical Considerations

Descriptive statistics were performed for all variables in both derivation and validation groups. Patient characteristics were summarized as counts with proportions for categorical variables and median with interquartile range for continuous variables. Results of the calculated probability based on the machine learning model have been presented in numbers ranging from 0 to 100, with 0 being the lowest probability of requiring intensive care, and 100 being the highest. The numbers were used to calculate area under the receiver operating characteristic curve (AUC) in the derivation and validation groups. For the derivation group, the mean value of the AUC for the 5 cross-validation sets of each model was used to compare the performance of the developed models. Receiver operating characteristic (ROC) curves were drawn, and the areas under the curves were calculated to assess the predictive performance of the models. P were calculated between the AUC of the machine learning model and the CURB-65 score. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-measures were measured for different cut-off values. Confusion matrices were constructed for both derivation and validation groups. All P values were two-sided, and a P value of <.05 was considered statistically significant. Statistical analysis was performed using R 4.0.0 (The R Foundation), with the pROC package to draw the ROC curves [19].

Results

Patient Characteristics

A total of 5193 patients with polymerase chain reaction–confirmed COVID-19 from 100 centers were registered with the nationwide cohort during the study period. Patients under 18 years (n=117, 2.2%) and those with missing data (n=289, 5.6%) were excluded, leaving 4787 patients for analysis. Among these patients, 3294 were assigned to the derivation group, and the remaining 1493 patients were assigned to the validation group (Figure 1).

Figure 1. Flowchart of the patient selection process. Patients hospitalized in 100 hospitals in South Korea from January 25, 2020, to June 3, 2020, were included. Patients who were admitted until March 20 were assigned to the derivation group, and those hospitalized after March 21 were assigned to the validation group.
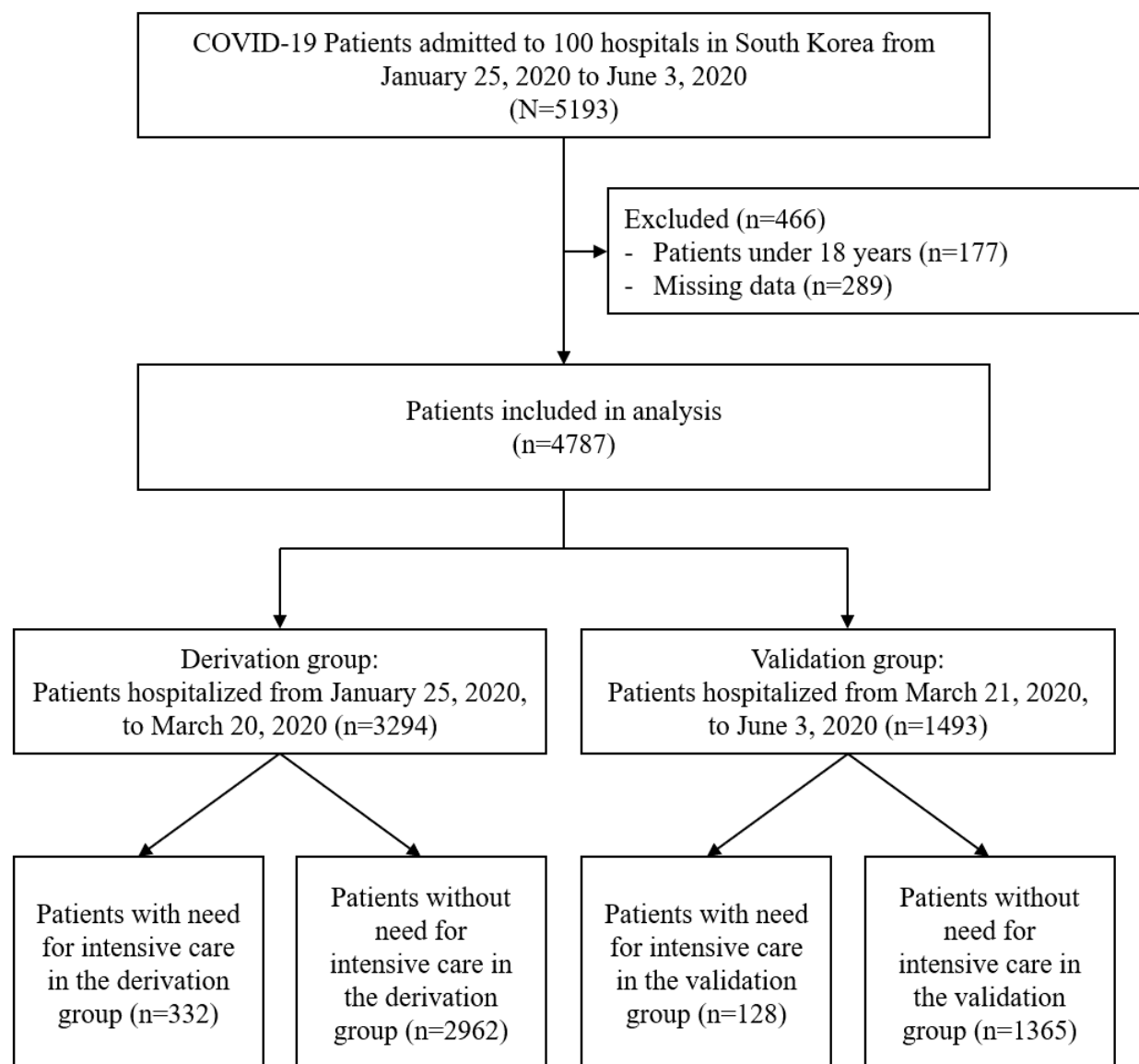
Compared to the patients in the validation group, those in the derivation group were older (median 57.0 years [IQR 42.0-68.0] vs 53.0 years [IQR 30.0-66.0]; *P*<.001), with a lower proportion of males (n=1227, 37.2% vs n=681, 45.6%; *P*<.001). Patients in the derivation group experienced more symptoms, including cough (n=1537, 46.6% vs n=440, 29.5%; *P*<.001), sputum (n=1054, 32.0% vs n=304, 20.4%; *P*<.001), headache (n=599, 18.2% vs n=165, 11.1%; *P*<.001), and myalgia (n=568, 17.2% vs n=159, 10.7%; *P*<.001), but less anosmia (n=40, 1.2% vs n=97, 6.5%; *P*<.001) compared to the validation group. The presence of underlying comorbidities was largely similar

between the two groups, except for hypertension (n=883, 26.8% vs n=351, 23.5%; *P*=.02) and diabetes (n=537, 16.3% vs n=204, 13.7%; *P*=.02), which were more common in the derivation group, and dementia (n=182, 5.5% vs n=153, 10.3%; *P*<.001), which was more common in the validation group. Patients in the derivation group were more independent in terms of their ADL compared to those in the validation group (n=2932, 89.0% vs n=1188, 79.6%; *P*<.001).

A total of 460 patients (9.6%) suffered the need for intensive care, of which 221 (4.6%) patients were admitted to the ICU, and 223 (4.7%) died (Table 1).

Table 1. Descriptive statistics for the included patients according to derivation and validation groups.

Variable	Total patients (N=4787)	Derivation group (n=3294)	Validation group (n=1493)	*P* value
Age (years), median (IQR)	55.0 (38.0-68.0)	57.0 (42.0-68.0)	53.0 (30.0-66.0)	<.001
Sex (male), n (%)	1908 (39.9)	1227 (37.2)	681 (45.6)	<.001
Smoking history, n (%)				<.001
Never smoked	4388 (91.7)	3084 (93.6)	1304 (87.4)	N/A[a]
Former smoker	136 (2.8)	97 (2.9)	39 (2.6)	N/A
Current smoker	263 (5.5)	114 (3.5)	149 (10.0)	N/A
Body temperature (°C), median (IQR)	36.8 (36.5-37.2)	36.9 (36.5-37.3)	36.8 (36.5-37.2)	.002
Symptoms, n (%)				
Cough	1977 (41.3)	1537 (46.6)	440 (29.5)	<.001
Sputum	1358 (28.4)	1054 (32.0)	304 (20.4)	<.001
Headache	764 (16.0)	599 (18.2)	165 (11.1)	<.001
Myalgia	727 (15.2)	568 (17.2)	159 (10.7)	<.001
Sore throat	688 (14.4)	513 (15.6)	175 (11.7)	.001
Dyspnea	654 (13.7)	543 (16.5)	111 (7.4)	<.001
Rhinorrhea	424 (8.9)	318 (9.7)	106 (7.1)	.005
Diarrhea	399 (8.3)	327 (9.9)	72 (4.8)	<.001
Chest pain	369 (7.7)	305 (9.3)	64 (4.3)	<.001
Nausea/vomiting	225 (4.7)	176 (5.3)	49 (3.3)	.002
Fatigue	188 (3.9)	149 (4.5)	39 (2.6)	.002
Anosmia	137 (2.9)	40 (1.2)	97 (6.5)	<.001
Hemoptysis	26 (0.5)	23 (0.7)	3 (0.2)	.051
Altered mentality	37 (0.8)	22 (0.7)	15 (1.0)	.29
Arthralgia	18 (0.4)	16 (0.5)	2 (0.1)	.11
Comorbidities, n (%)				
Hypertension	1234 (25.8)	883 (26.8)	351 (23.5)	.02
Diabetes	741 (15.5)	537 (16.3)	204 (13.7)	.02
Dementia	335 (7.0)	182 (5.5)	153 (10.3)	<.001
Chronic cardiac disease	195 (4.1)	142 (4.3)	53 (3.6)	.25
Cancer	160 (3.3)	113 (3.4)	47 (3.2)	.68
Asthma	123 (2.6)	95 (2.9)	28 (1.9)	.052
Chronic liver disease	80 (1.7)	56 (1.7)	24 (1.6)	.92
Heart failure	70 (1.5)	44 (1.3)	26 (1.7)	.34
Chronic kidney disease	60 (1.3)	48 (1.5)	12 (0.8)	.08
Chronic obstructive pulmonary disease	42 (0.9)	35 (1.1)	7 (0.5)	.06
Chronic neurologic disorder	42 (0.9)	24 (0.7)	18 (1.2)	.14
Chronic hematologic disorder	35 (0.7)	28 (0.8)	7 (0.5)	.21
Autoimmune disease	34 (0.7)	27 (0.8)	7 (0.5)	.25
Pregnancy	20 (0.4)	13 (0.4)	7 (0.5)	.90
HIV infection	10 (0.2)	7 (0.2)	3 (0.2)	>.99
Activities of daily living, n (%)				<.001
Independent	4120 (86.1)	2932 (89.0)	1188 (79.6)	N/A
Partially dependent	375 (7.8)	203 (6.2)	172 (11.5)	N/A

Variable	Total patients (N=4787)	Derivation group (n=3294)	Validation group (n=1493)	P value
Totally dependent	292 (6.1)	160 (4.9)	132 (8.8)	N/A
Need for intensive care, n (%)	460 (9.6)	332 (10.1)	128 (8.6)	.12
Death	223 (4.7)	161 (4.9)	62 (4.2)	.30
Admission to ICU[b]	221 (4.6)	169 (5.1)	52 (3.5)	.02
Vasopressor treatment	119 (2.5)	84 (2.5)	35 (2.3)	.75
Mechanical ventilation	66 (1.4)	54 (1.6)	12 (0.8)	.001
Extracorporeal life support	27 (0.6)	21 (0.6)	6 (0.4)	.43

[a]N/A: not applicable.

[b]ICU: intensive care unit.

Derivation and Internal Validation of the Machine Learning Model

With the AutoML, 55 machine learning models were developed to predict the need for intensive care among patients with COVID-19. The XGBoost model, which showed an AUC of 0.897 (95% CI 0.877-0.917) by cross-validation in the derivation group, was chosen as the best machine learning model (Multimedia Appendix 1). The important features of this model were ADL, age, dyspnea, initial body temperature, sex, and underlying comorbidities. More detailed information on each feature is presented in Multimedia Appendix 2. The developed machine learning model revealed significantly better discrimination performance than the CURB-65 score (AUC 0.836 with 95% CI 0.825-0.847, P<.001) for predicting the need for intensive care among patients with COVID-19. A comparison of the ROC curves for the XGBoost machine learning model and the CURB-65 score is shown in Figure 2A.

Figure 2. Receiver operating characteristic curves for the machine learning model (XGBoost) and the CURB-65 (confusion, urea, respiratory rate, blood pressure, and 65 years of age or older) score for predicting patients requiring intensive care. (A) Comparison in the derivation group, where the area under the receiver operating characteristic (AUC) curves were 0.897 for the gradient boosting machine model, and 0.836 for the CURB-65 score (P<.001). (B) Comparison in the temporal external validation group, where the AUC were 0.885 for the machine learning model, and 0.843 for the CURB-65 score (P=.01).
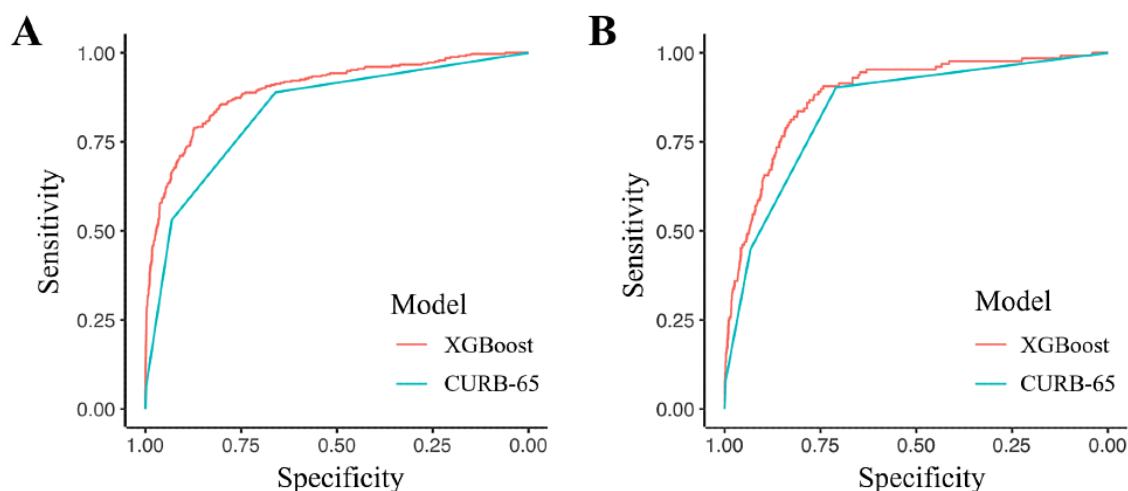
External Validation of the Model

External validation was performed in the validation group, using the developed XGBoost machine learning model. The discrimination performance of the machine learning model showed an AUC of 0.885 with 95% CI 0.855-0.915, which was significantly higher than that of CURB-65 (0.843, 95% CI 0.829-0.857, P=.01) (Figure 2B).

Comparison of the Machine Learning Model With the CURB-65 Score With Different Thresholds

With a cut-off value of 0.5, the CURB-65 score showed a sensitivity of 0.89, specificity of 0.66, PPV of 0.05, NPV of 1.00, and F-measure of 0.10. A cut-off value of 0.06 for the XGBoost machine learning model, which shows a similar sensitivity (0.89), revealed a specificity of 0.75, PPV of 0.36, NPV of 0.99, and F-measure of 0.43. The XGBoost score also revealed better specificity, PPV, and F-measure compared to CURB-65 when different cut-off thresholds were used (Table 2). The confusion matrices of the developed model for the development and validation groups are shown in Multimedia Appendix 3.

Table 2. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-measure for the machine learning model and the CURB-65 (confusion, urea, respiratory rate, blood pressure, and 65 years of age or older) score, with different cut-offs.

Cut-off	Sensitivity	Specificity	PPV	NPV	F-measure
CURB-65 score >0.5	0.89	0.66	0.05	1.00	0.10
XGBoost score >0.06	0.89	0.75	0.36	0.99	0.43
CURB-65 score >1.5	0.53	0.93	0.14	0.99	0.22
XGBoost score >0.34	0.53	0.97	0.63	0.95	0.58
CURB-65 score >2.5	0.06	1.00	0.40	0.98	0.11
XGBoost score >0.89	0.06	1.00	0.95	0.90	0.12

Web Application of Prediction Models

A web-based application was developed for better accessibility and easy use of the models. The application can be accessed online [20] (Figure 3), and it has been enlisted in the World Health Organization's Digital Health Atlas [21]. The application calculates the probability of need for intensive care, which is computed according to the derived model. However, it does not store any data yet. It is intended for use by medical practitioners to aid with medical decisions.

Figure 3. Screenshots of the web-based application for easy usage of the developed machine learning model [20]. After input of simple patient-derived information, the probability of the need for intensive care within 30 days is calculated.
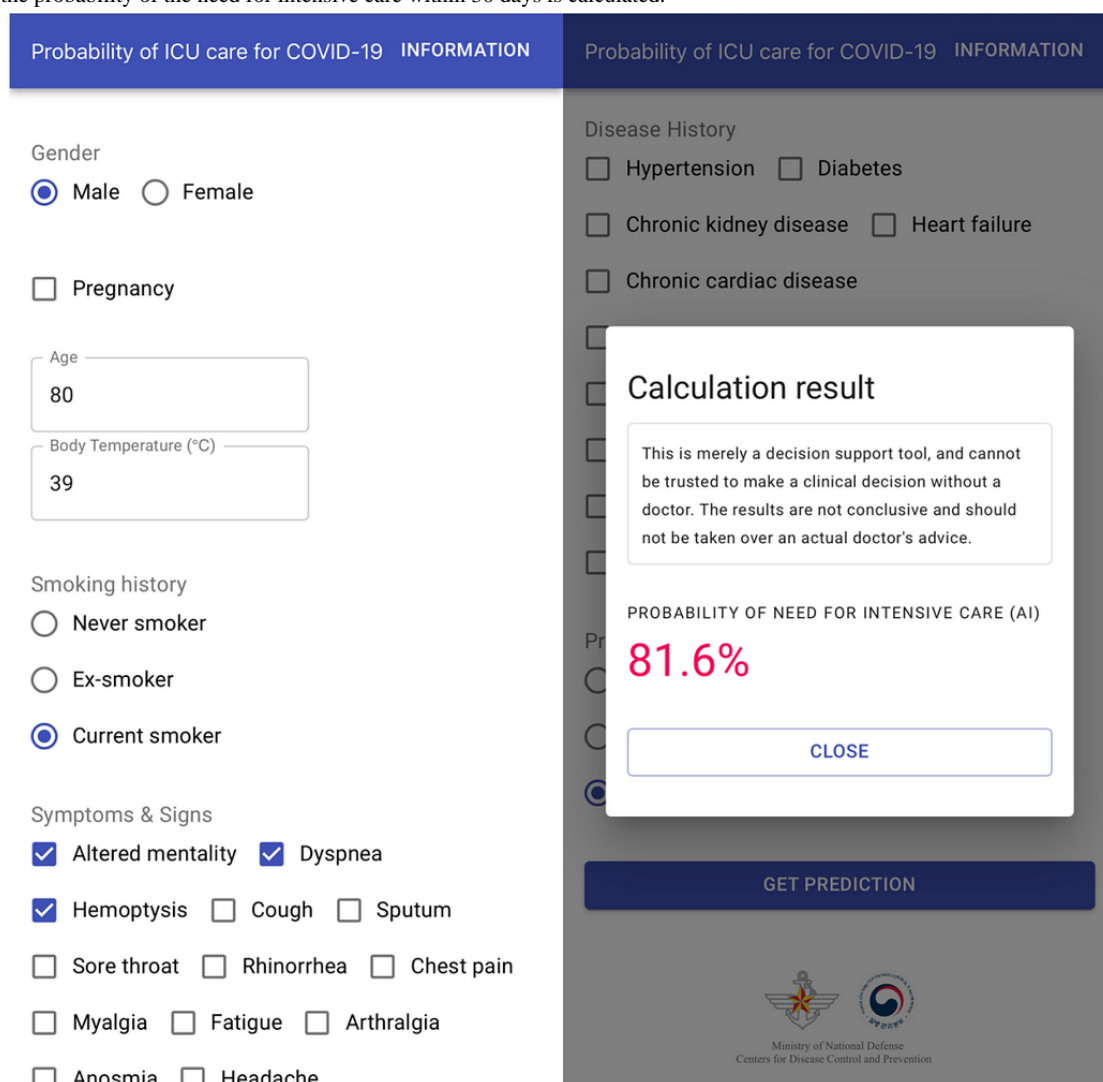
Discussion

Principal Results

This study presents a machine learning model that predicts the need for intensive care among patients with COVID-19 from a nationwide cohort in South Korea, including 100 hospitals. The model was derived from the data of patients who were hospitalized from January 25, 2020, to March 20, 2020, and was validated in a separate group of patients hospitalized between March 21, 2020, and June 3, 2020. The AUC of the machine learning model was 0.897 (95% CI 0.877-0.917) for the derivation cohort and 0.885 (95% CI 0.855-0.915) for the validation cohort, which revealed better discrimination performance than that of CURB-65. Important features included ADL, age, dyspnea, initial body temperature, and sex.

Comparison With Prior Work

The main features selected in the machine learning model are mostly coherent with previous reports. Older age and male sex have been constantly emphasized as major risk factors for adverse outcomes in patients with COVID-19 [6,22,23]. An early report on 85 fatal cases of COVID-19 in Wuhan [24] revealed that the mean age of patients was 65.8 years, and 62 of the 85 patients (72.9%) were male [5]. Dyspnea was also a major factor in our study. Incidence of dyspnea is relatively low in COVID-19 as compared to other respiratory symptoms, despite the common pneumonic infiltration on chest radiographs [25]. In a recent systematic review that included 43 studies [3], shortness of breath was observed in 49.2% in patients with critical illness, while bilateral pneumonia was observed in chest computed tomography (CT) images of 91.0% of patients with the same disease extent. In an earlier study in China, even in severely ill patients, dyspnea was observed in about 37.6% of the patients [26]. Therefore, the presentation of dyspnea may imply extensive involvement of the lungs, which leads to grave prognoses [5]. Underlying comorbidities were also repeatedly highlighted as major risk factors for poor prognoses of patients with COVID-19. A pooled analysis of COVID-19 reports emphasized that hypertension is associated with an approximately 2.5-fold increased risk of higher severity and mortality [27]. Another previous study of 174 patients revealed that patients with diabetes were at higher risk of pneumonia, release of tissue injury-related enzymes, and higher rates of inflammatory responses [28]. Such findings are well summarized in a systematic review that included 3027 patients [5]: male sex (pooled odds ratio [OR] 1.76 with 95% CI 1.41-2.18), age over 65 years (pooled OR 6.06 with 95% CI 3.98-9.22), dyspnea (pooled OR 4.16 with 95% CI 3.13-5.53), presence of cardiovascular disease (pooled OR 5.19 with 95% CI 3.25-8.29), diabetes (pooled OR 3.68 with 95% CI 2.68-5.03), and hypertension (pooled OR 2.72 with 95% CI 1.60-4.64) were all significant factors associated with the progression of COVID-19.

In addition to previous reports, ADL limitation and abnormal body temperature were associated with the need for intensive care among patients with COVID-19 in our study. ADL limitation is known to be an independent risk factor for mortality among elderly patients with pneumonia [29,30]. Because most of the poor outcomes occur in the elderly in COVID-19 [6,22,23], it is probable that ADL limitation leads to the need for intensive care. Abnormal body temperature is also a well-known risk factor for grave prognosis in community-acquired pneumonia patients [31].

Strengths of This Study

Our machine learning prediction model based on simple patient demographics and subjective symptoms can be useful for the early triage of patients in this pandemic situation. First, it uses information that can be easily provided without advanced equipment, such as age, sex, past medical history, and subjective symptoms. Previous scoring systems [10,11,32-35], including a recently reported deep learning model [36], require laboratory or radiographic findings as the main variables. Although such models can be helpful in fully equipped medical facilities, they initially consume a certain amount of medical resources and time. In areas where laboratory exams or CT exams are limited, our scoring model can be an effective solution for earlier triage. Second, because our model is based on the AutoML technique [20], the relative importance of the features can easily be adjusted with the newly acquired patient data. AutoML techniques have been studied extensively [14] and are expected to be useful for many applications, including in the field of health care [17]. AutoML mainly helps in building machine learning pipelines, which requires expertise in machine learning and is time consuming. It is effective when the time or resources necessary for building a high-functioning model are limited. Considering the rapid adaptability of our model, it can be used effectively with populations with different ethnic or regional backgrounds when further data are collected from the similar populations. It is useful in this pandemic situation, where insufficiency of medical resources has been identified as a critical factor in patient survival [7]. Especially in contexts with less than adequate medical staff, the web-based application is easy to use owing to its intuitive interfaces and clear guides, making it possible for the attending physicians to triage patients without adequate medical knowledge about COVID-19.

CURB-65 was used for comparison with the AutoML model in our study. CURB-65 is a well-known score derived and validated for predicting mortality among patients with community-acquired pneumonia [16], and also shows promising performance in patients with COVID-19 [37]. It is comprised of 6 variables: mental status, levels of blood urea nitrogen, respiratory rate, blood pressure, and age. COVID-19 commonly accompanies pneumonia [4]. In a recent systematic review [38], bilateral (72.9%) or unilateral (25.0%) involvement of chest X-rays was observed among patients with confirmed COVID-19. A large proportion of the involvement is ground-glass opacities (68.5%), which are difficult to recognize from simple chest radiographs. In our study, 2050 of the 4787 patients (42.8%) underwent chest CT evaluation, and among them, 1535 (74.9%) were recognized to have pneumonic infiltrations.

Recommendations

Our model can be used as a decision-support system for medical professionals when active monitoring is not possible due to patient overload caused by the lack of availability of medical staff. However, we cannot recommend a uniform cut-off value for patient transfer to higher-level facilities because this decision

depends on the local situation. This decision needs to be made considering the availability of beds in higher-level facilities, the rate of regional increase in the number of patients with COVID-19, and the treatment capability of the facility the patient is currently admitted to. Yet, one solid recommendation that can be made is to prioritize the transfer of patients with a higher probability of need for intensive care when feasible.

Limitations

Our study had several limitations. First, the sample excluded patients assigned to community treatment centers. However, assignment to community treatment centers was mostly conducted for quarantining purposes, not for active treatment. When they required active treatment, such patients were transferred to hospitals and were eventually included in this study. Second, our data set was imbalanced, with 9.6% of the patients requiring intensive care. Third, our initial model was built based on patients from South Korea. Nevertheless, due to the nature of AutoML, the model can be updated easily when further data become available.

Conclusions

In conclusion, we derived and validated a machine learning prediction model comprising simple patient-provided characteristics. The model included variables that were largely consistent with previous reports, and it can efficiently anticipate deterioration among patients with COVID-19. The model is easy to use and adjust, requires minimal resources, and can be an effective solution for easy triage in areas with a shortage of medical resources. The model can be used for patient monitoring, and also has a potential as a warning system for self-quarantined patients. However, in the future, randomized trials need to be conducted to examine the direct impact of our model on patient survival.

Authors' Contributions

HJK, DH, and JH contributed to the conception and design of the study. DH, JHK, DK, BH, WS, YKL, DL, SOH, MJP, and JH were responsible for data collection. HJK, DH, and JH contributed to the analysis and interpretation of data. HJK and JH drafted the paper. DH, JHK, DK, BH, YKL, DL, SOH, MJP, and WS revised the paper for important intellectual content. HJK, DH, JHK, DK, BH, WS, YKL, DL, SOH, MJP, and JH approved the final version of the manuscript to be submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
AUROC curves of the developed machine learning models.
[DOCX File , 19 KB - jmir_v22i11e24225_app1.docx]

Multimedia Appendix 2
Feature importance of the gradient boosting machine model for prediction of patients requiring intensive care.
[DOCX File , 16 KB - jmir_v22i11e24225_app2.docx]

Multimedia Appendix 3
Confusion matrices for the development and validation groups.
[DOCX File , 16 KB - jmir_v22i11e24225_app3.docx]

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020 Aug 31. URL: https://covid19.who.int/ [accessed 2020-09-01]
2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2) [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
3. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
4. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]

XSL•FO
RenderX

5. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. J Infect 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.021] [Medline: 32335169]

6. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar 28;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]

7. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. The Lancet Global Health 2020 Apr;8(4):e480 [FREE Full text] [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]

8. Zhang Z, Yao W, Wang Y, Long C, Fu X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. J Infect 2020 Jul;81(1):147-178 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.018] [Medline: 32209384]

9. Shim E, Mizumoto K, Choi W, Chowell G. Estimating the Risk of COVID-19 Death During the Course of the Outbreak in Korea, February-May 2020. J Clin Med 2020 May 29;9(6) [FREE Full text] [doi: 10.3390/jcm9061641] [Medline: 32485871]

10. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]

11. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. Clin Infect Dis 2020 Apr 09 [FREE Full text] [doi: 10.1093/cid/ciaa414] [Medline: 32271369]

12. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]

13. Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. Radiology 2020 Aug;296(2):E86-E96 [FREE Full text] [doi: 10.1148/radiol.2020201433] [Medline: 32301647]

14. He X, Zhao K, Chu X. AutoML: A Survey of the State-of-the-Art. arXiv. [FREE Full text] Preprint posted online Aug 2, 2019.

15. iCReaT. National Institute of Health. URL: https://icreat.nih.go.kr/ [accessed 2020-09-04]

16. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003 May;58(5):377-382 [FREE Full text] [doi: 10.1136/thorax.58.5.377] [Medline: 12728155]

17. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 2020 Apr;104:101822 [FREE Full text] [doi: 10.1016/j.artmed.2020.101822] [Medline: 32499001]

18. AutoML: Automatic Machine Learning. H2O.ai. 2020. URL: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html [accessed 2020-09-30]

19. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 Mar 17;12:77 [FREE Full text] [doi: 10.1186/1471-2105-12-77] [Medline: 21414208]

20. Probability of ICU care for COVID-19. Ministry of National Defence, Centers for Disease Control and Prevention. URL: http://icu.docl.org [accessed 2020-09-04]

21. DOCL Outcome Prediction Platform (Doctors on the Cloud). Digital Health Atlas. URL: https://digitalatlas.who.int/en/covid-19/projects/1274/published [accessed 2020-09-01]

22. Bhargava A, Fukushima EA, Levine M, Zhao W, Tanveer F, Szpunar SM, et al. Predictors for Severe COVID-19 Infection. Clin Infect Dis 2020 May 30 [FREE Full text] [doi: 10.1093/cid/ciaa674] [Medline: 32472676]

23. Du R, Liang L, Yang C, Wang W, Cao T, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. Eur Respir J 2020 May;55(5) [FREE Full text] [doi: 10.1183/13993003.00524-2020] [Medline: 32269088]

24. Du Y, Tu L, Zhu P, Mu M, Wang R, Yang P, et al. Clinical Features of 85 Fatal Cases of COVID-19 from Wuhan. A Retrospective Observational Study. Am J Respir Crit Care Med 2020 Jun 01;201(11):1372-1379 [FREE Full text] [doi: 10.1164/rccm.202003-0543OC] [Medline: 32242738]

25. Bertran Recasens B, Martinez-Llorens JM, Rodriguez-Sevilla JJ, Rubio MA. Lack of dyspnea in patients with Covid-19: another neurological conundrum? Eur J Neurol 2020 Apr 17 [FREE Full text] [doi: 10.1111/ene.14265] [Medline: 32301553]

26. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]

27. Lippi G, Wong J, Henry BM. Hypertension in patients with coronavirus disease 2019 (COVID-19): a pooled analysis. Pol Arch Intern Med 2020 Apr 30;130(4):304-309 [FREE Full text] [doi: 10.20452/pamw.15272] [Medline: 32231171]

28. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev 2020 Mar 31:e3319 [FREE Full text] [doi: 10.1002/dmrr.3319] [Medline: 32233013]

29. Torres OH, Muñoz J, Ruiz D, Ris J, Gich I, Coma E, et al. Outcome predictors of pneumonia in elderly patients: importance of functional assessment. J Am Geriatr Soc 2004 Oct;52(10):1603-1609. [doi: 10.1111/j.1532-5415.2004.52492.x] [Medline: 15450034]

30. Mody L, Sun R, Bradley SF. Assessment of pneumonia in older adults: effect of functional status. J Am Geriatr Soc 2006 Jul;54(7):1062-1067 [FREE Full text] [doi: 10.1111/j.1532-5415.2006.00797.x] [Medline: 16866676]

31. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 1997 Jan 23;336(4):243-250. [doi: 10.1056/NEJM199701233360402] [Medline: 8995086]

32. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J, et al. Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019. Crit Care Med 2020 Aug 01;48(8):e657-e665 [FREE Full text] [doi: 10.1097/CCM.0000000000004411] [Medline: 32697506]

33. Mahdjoub E, Mohammad W, Lefevre T, Debray M, Khalil A, Study Group§. Admission chest CT score predicts 5-day outcome in patients with COVID-19. Intensive Care Med 2020 Aug;46(8):1648-1650. [doi: 10.1007/s00134-020-06118-y] [Medline: 32468085]

34. Galloway JB, Norton S, Barker RD, Brookes A, Carey I, Clarke BD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study. J Infect 2020 Aug;81(2):282-288 [FREE Full text] [doi: 10.1016/j.jinf.2020.05.064] [Medline: 32479771]

35. Sabri A, Davarpanah AH, Mahdavi A, Abrishami A, Khazaei M, Heydari S, et al. Novel coronavirus disease 2019: predicting prognosis with a computed tomography-based disease severity score and clinical laboratory data. Pol Arch Intern Med 2020 Aug 27;130(7-8):629-634 [FREE Full text] [doi: 10.20452/pamw.15422] [Medline: 32500700]

36. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun 2020 Jul 15;11(1):3543 [FREE Full text] [doi: 10.1038/s41467-020-17280-8] [Medline: 32669540]

37. Satici C, Demirkol MA, Sargin Altunok E, Gursoy B, Alkan M, Kamat S, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. Int J Infect Dis 2020 Sep;98:84-89 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.038] [Medline: 32553714]

38. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020 Mar 13:101623 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101623] [Medline: 32179124]

Abbreviations

ADL: activities of daily living
AUC: area under the receiver operating characteristic curve
AutoML: automated machine learning
CT: computed tomography
CURB-65: confusion, urea, respiratory rate, blood pressure, and 65 years of age or older
ICU: intensive care unit
NPV: negative predictive value
OR: odds ratio
PPV: positive predictive value
ROC: receiver operating characteristic

XSL•FO

RenderX

Original Paper

Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study

Andreas Älgå[1,2], MD, PhD; Oskar Eriksson[3], BSc; Martin Nordberg[1], MD, PhD

[1]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

[2]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

[3]DataRobot Inc, Stockholm, Sweden

Corresponding Author:
Andreas Älgå, MD, PhD
Department of Clinical Science and Education, Södersjukhuset
Karolinska Institutet
Sjukhusbacken 10
Stockholm, 118 83
Sweden
Phone: 46 8 616 10 00
Email: andreas.alga@ki.se

Related Articles:

Comment in: https://www.jmir.org/2021/5/e27937

Comment in: https://www.jmir.org/2021/5/e29156

Abstract

Background: The COVID-19 pandemic has spread at an alarming speed, and effective treatment for the disease is still lacking. The body of evidence on COVID-19 has been increasing at an impressive pace, creating the need for a method to rapidly assess the current knowledge and identify key information. Gold standard methods such as systematic reviews and meta-analyses are regarded unsuitable because they have a narrow scope and are very time consuming.

Objective: This study aimed to explore the published scientific literature on COVID-19 and map the research evolution during the early phase of the COVID-19 pandemic.

Methods: We performed a PubMed search to analyze the titles, keywords, and abstracts of published papers on COVID-19. We used latent Dirichlet allocation modeling to extract topics and conducted a trend analysis to understand the temporal changes in research for each topic, journal impact factor (JIF), and geographic origin.

Results: Based on our search, we identified 16,670 relevant articles dated between February 14, 2020, and June 1, 2020. Of these, 6 articles were reports from peer-reviewed randomized trials on patients with COVID-19. We identified 14 main research topics, of which the most common topics were health care responses (2812/16,670, 16.86%) and clinical manifestations (1828/16,670, 10.91%). We found an increasing trend for research on clinical manifestations and protective measures and a decreasing trend for research on disease transmission, epidemiology, health care response, and radiology. Publications on protective measures, immunology, and clinical manifestations were associated with the highest JIF. The overall median JIF was 3.7 (IQR 2.6-5.9), and we found that the JIF for these publications declined over time. The top countries producing research were the United States, China, Italy, and the United Kingdom.

Conclusions: In less than 6 months since the novel coronavirus was first detected, a remarkably high number of research articles on COVID-19 have been published. Here, we discuss and present the temporal changes in the available COVID-19 research during the early phase of the pandemic. Our findings may aid researchers and policy makers to form a structured view of the current COVID-19 evidence base and provide further research directions.

KEYWORDS

COVID-19; SARS-CoV-2; coronavirus; pandemic; topic modeling; research; literature

Introduction

The novel coronavirus (SARS-CoV-2) was first detected in the Hubei Province in China in December 2019 [1]. The virus is known to cause a severe respiratory disease (COVID-19) that has rapidly spread worldwide [2]. On March 11, 2020, the World Health Organization declared the COVID-19 outbreak a pandemic [3]. Owing to the novelty of the disease, its clinical course and treatment are largely unknown [4]. However, the scientific society has mobilized quickly, and by June 1, 2020, more than 1300 clinical trials had been registered at various clinical trial registry sites [5,6].

Keeping track of the growing evidence base in medicine is becoming increasingly difficult owing to a large number of publications [7]. A rapid assessment of a dynamic research field such as COVID-19, where the body of evidence has been increasing at an impressive pace, requires an approach that is more direct and has a wider scope than that of the current gold standard methods, such as scoping and systematic reviews [8]. The potential uses of machine learning and artificial intelligence in the fight against the COVID-19 crisis has been previously discussed [9].

Several systematic reviews have been published on specific aspects of the pandemic, such as the impact of comorbidities, symptoms, and treatments [10-12]. However, a comprehensive assessment of all the available scientific publications on COVID-19 is lacking. Therefore, we aimed to explore the published scientific literature on COVID-19, assess relevant topics, and map the research evolution during the early phase of the COVID-19 pandemic by using a machine learning–based approach.

Methods

Data Collection

We searched PubMed on June 1, 2020, using PubMed E-utilities [13] and the search terms "covid[Title/Abstract] OR covid-19[Title/Abstract]", with no language or date restrictions, utilizing the Biopython package for Python 3.6 (Python Software Foundation). The focus of this study was the disease (COVID-19) as opposed to the virus; hence, we refrained from including search terms such as "coronavirus". We retrieved the title, keywords, abstract, date of last revision, list of author affiliation, journal name, and PubMed identifier number for each publication. We used the date of last revision as the indexing date for all articles, as this date was the readily available for all the data, unlike the date of submission or publishing.

Data Preprocessing

All text data were lowercased and cleaned for double spaces, special characters, and numbers. Subsequently, we applied a list of stop words from PubMed [14], general English stop words, and subject-specific stop words (eg, covid, corona) to the titles and abstracts to remove non–information-bearing words from the text (Multimedia Appendix 1). The data processing workflow used is depicted in Figure 1.

Figure 1. Data processing flowchart. Italicized text indicate the method used. JIF: journal impact factor.
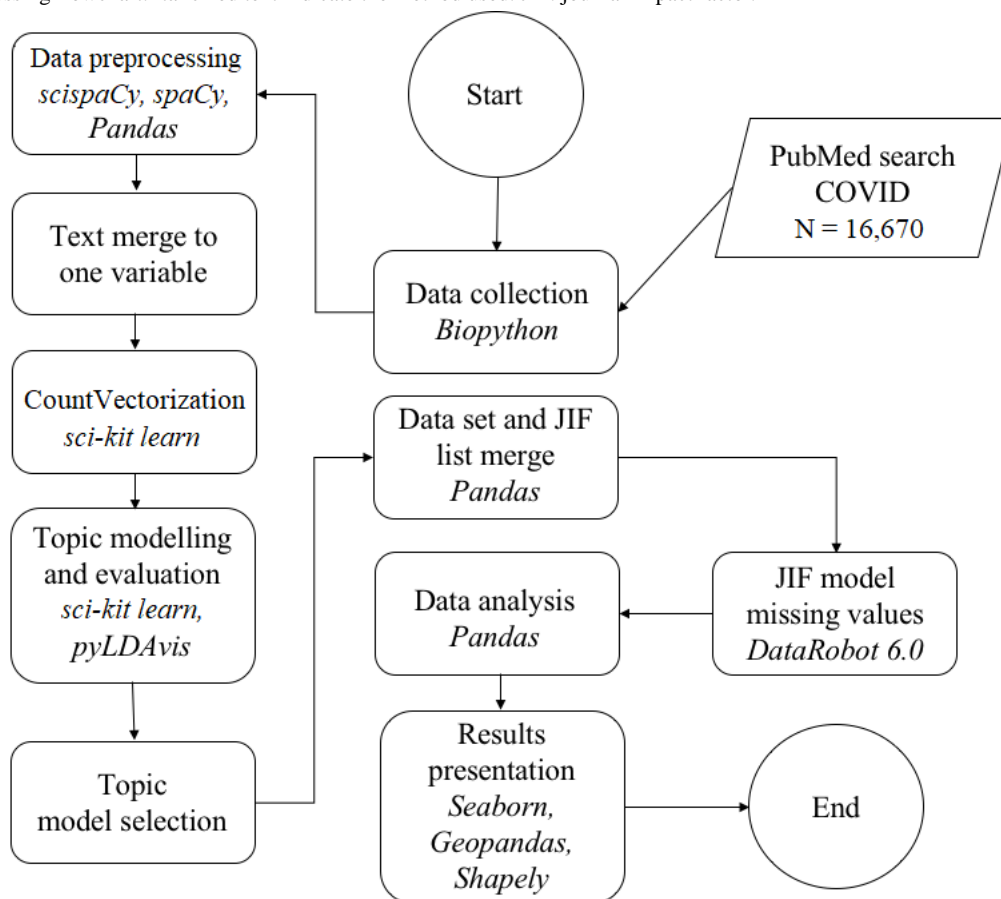
We lemmatized the text data using the Python package scispaCy. We considered the country of the first author's affiliation to be the country of origin and extracted geographic entities from the affiliations using the Python package spaCy. If a country name was not included in the affiliation, we used the last geographic entity mentioned and manually linked this geographic entity to a country; for instance, "New York" was linked to "the United States."

To identify randomized clinical trials, we searched for the words "randomis*" and "randomiz*" in the titles. We then manually assessed all articles identified as potential randomized clinical trials to determine their true article type.

Latent Dirichlet Allocation

Latent Dirichlet allocation (LDA) is a generative statistical model for data collection [15] that has previously shown to be suitable for topic modeling in medicine [16,17]. We used LDA to derive useful information from the identified articles. We concentrated all the textual data collected from each article (ie, title, keywords, and abstracts when available in PubMed) into one variable. We then used this variable as the text corpus for the whole data set and subsequently vectorized the variable using CountVectorizer in the Python package scikit-learn.

To assess the topic spaces in an interspersed arbitrary selection of topic numbers, we computed LDA models and principal component analysis (PCA) plots for 8, 13, and 35 topics. After assessing the results, we subsequently computed LDA models for all numbers of topics between 3 and 23. We chose the numbers 3 and 23 because fewer than 3 topics were considered too few for any fathomable use case and more than 23 topics were considered not useful or comprehensible, based on expert opinion. We recorded the evaluation metrics for perplexity, leave-out likelihood, and graphical PCA for each model. We decided the final number of topics based on the assessments of these three evaluation metrics, as well as the authors' domain knowledge of COVID-19 and medical research.

Thereafter, 6 experienced clinicians and researchers independently labeled the identified topics based on the 15 most-common keywords for the articles assigned to each topic and the resulting most-frequent words in the LDA model for each topic (Multimedia Appendix 2). We then discussed the proposed labels until a consensus for each topic label was reached. We performed LDA modeling using the Python package scikit-learn and plotted the results using the package pyLDAvis [18].

We identified the most probable topic of each article and assigned it as the articles' main topic. The weekly number of articles for each topic was then computed and the time series for the proportion of each topic during the week were plotted for further analysis.

Journal Impact Factor

We looked up the journal names of all articles obtained from our PubMed search against the 2019 list of journal impact factor (JIF; Journal Citations Report, Clarivate Analytics), which covered 12,515 scientific journals [19]. Journal names were

matched using DataRobot 6.0 (DataRobot Inc.) fuzzy matching, wherein a 90% similarity was considered a match. We then manually corrected the list of matched journal names to avoid mismatch. In cases where the journal name of the article could not be matched against the Clarivate JIF list, we developed a random forest regression model in DataRobot to predict the JIF from the article data.

DataRobot automatically performs a modelling competition in which a wide selection of algorithm and data preprocessing techniques compete with one another. The model with the best root mean square error (RMSE) and R^2 on hold-out data is then chosen as the champion model. A modelling competition is beneficial because the same algorithm can have different efficiencies on different data sets. The article data that we could not match with the Clarivate JIF list originated from the journals that were not identified by the trained model. Therefore, we excluded all journal-specific variables when using the model, to avoid overfitting the model to the training data. To evaluate the model, we used a test sample and RMSE and R^2 as model performance metrics. In the model, we used affiliations of the authors; keywords; first author's affiliation; and all text in the title, keywords, and abstract—together and as separate variables. We built and deployed the random forest regression model in DataRobot.

Subsequently, we computed the median JIF for each identified topic. We also calculated the median JIF over time and the median JIF in correlation to the number of articles.

Data Analysis

We performed all data analyses and visualizations using the Python packages Panda, NumPy, Matplotlib, pyLDAvis, and Seaborn (Python Software Foundation).

Results

Search Results

We found 16,670 articles dated between February 14, 2020, and June 1, 2020. Of these 16,670 articles, abstracts were available for 8560 (51.34%) articles, whereas title and keywords were available for 16,623 (99.71%) and 7739 (46.42%) articles, respectively.

The median weekly number of articles published was 374.5 (IQR 29.0-1648.5), ranging from 2 during week 7 (February 10-16, 2020) to 4536 during week 22 (May 25-30, 2020). The number of articles published per week shows an exponential development over time (Figure 2). The number of published case reports and review articles started to increase 8 and 11 weeks, respectively, after the first article was published. The number of published reports from randomized clinical trials or protocols for randomized clinical trials were uniformly spread throughout the study period. We conducted a manual search of the 23 articles containing either "randomiz*" or "randomis*" in the title and found that 6 (26%) of those were actual reports from randomized clinical trials.

Figure 2. Number of articles published per week and the corresponding median journal impact factor.
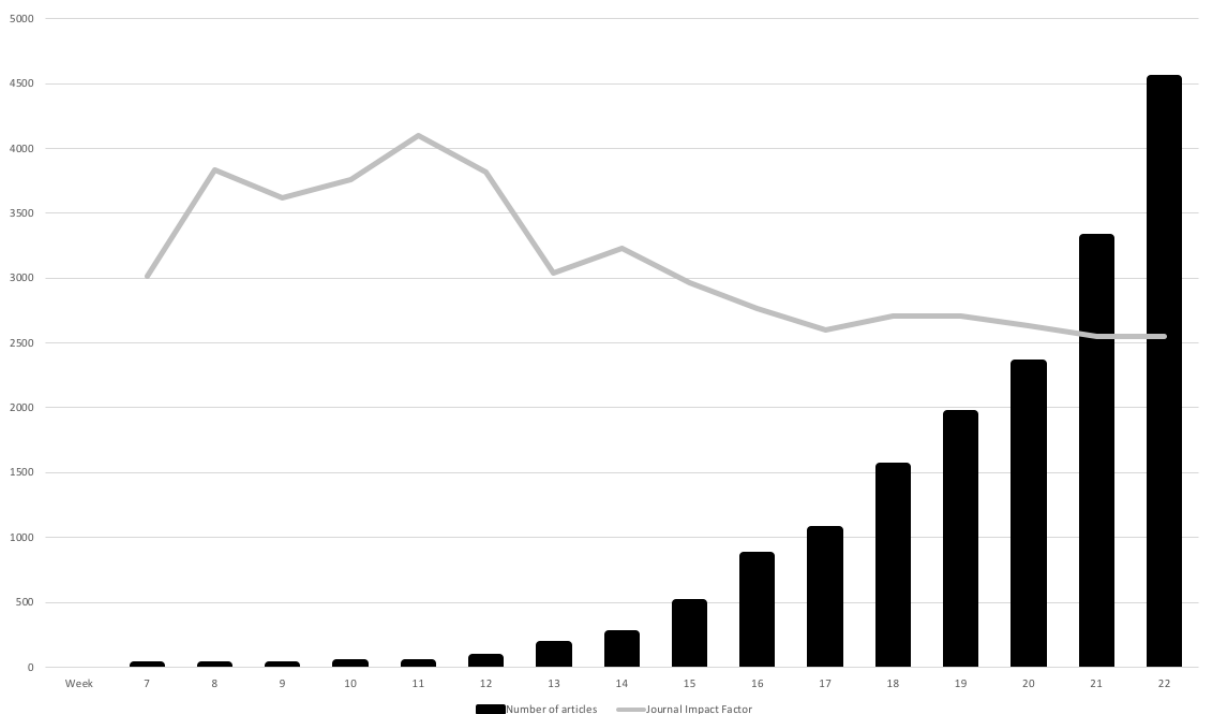
Analysis Based on LDA Modeling and Topics

The evaluation metrics perplexity and leave-out likelihood for LDA models with 3-23 topics (ie, the numbers of topics chosen based on a combination of expert opinion and arbitrary selection of 3 topics) exhibited decreasing and increasing values through

the whole set; hence, these metrics provided little additional value with regard to choosing the optimal number of topics (Multimedia Appendix 3). In contrast, by assessing the PCA plots for all 20 models, we found the optimal number of topics (ie, the number of solutions with the least amount of overlapping topics) to be 14 (Figure 3). We found a low correlation between

topic occurrence in the same article, indicating articles had   well-defined topics ([Multimedia Appendix 4](#)).

Figure 3. Principal component analysis plot for the chosen latent Dirichlet allocation model with 14 topics. Overlaps are seen for three topic clusters; however, these topics were found to be separated by clinical relevance.
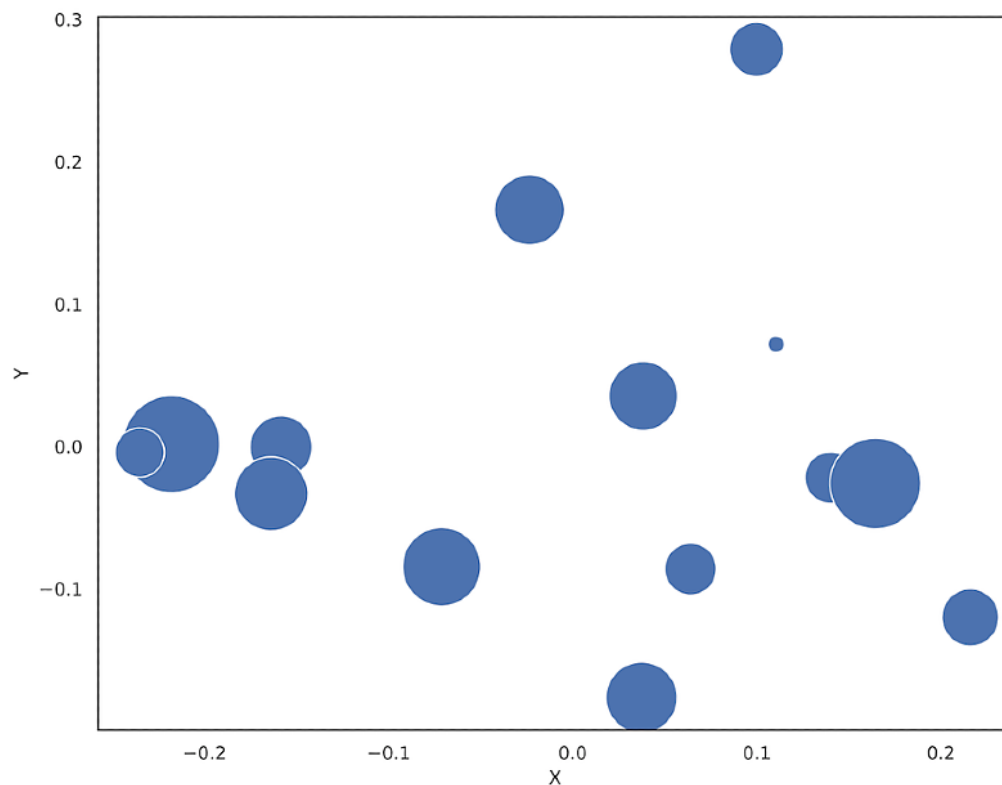
The Pearson correlation coefficients ranged from 0.00 to 0.17, where the correlation between the topics health care response and clinical manifestations was the highest. [Table 1](#) shows the labels, corresponding 5 most-frequent PubMed keywords and terms based on LDA, and the number of articles published for each topic. The main topics were health care response, with 2812 (16.86%) publications; clinical manifestations, with 1828 (10.96%) publications; and psychological impact, with 1771 (10.62%) publications. The least common topics were risk factors, laboratory diagnostics, and immunology.

Table 1. COVID-19 topics from latent Dirichlet allocation modeling.

Topic No.	Label	Five most-frequent terms based on latent Dirichlet allocation	Five most-frequent PubMed keywords	Journal impact factor, median (IQR)	Articles published, n (%) (N=16,670)
1	Laboratory diagnostics	antibody, time, laboratory, diagnostic, assay	pcr, rt, testing, disease, test	3.36 (2.0-6.1)	599 (3.59)
2	Therapies and vaccines	chloroquine, anti, hydroxychloroquine, pandemic, potential	hydroxychloroquine, chloroquine, drug, disease, antiviral	4.10 (2.9-6.6)	1193 (7.15)
3	Risk factors	ecmo, renin, respiratory, clinical, risk	diabetes, angiotensin, ace, disease, enzyme	4.13 (2.9-6.5)	420 (2.51)
4	Health care response	worker, response, practice, service, recommendation	health, pandemic, public, infection, disease	3.39 (2.4-5.1)	2812 (16.86)
5	Epidemiology	risk, control, datum, period, rate	disease, respiratory, epidemiology, novel, infection	4.09 (2.8-6.3)	819 (4.91)
6	Disease transmission	cause, spread, health, transmission, outbreak	respiratory, disease, syndrome, acute, virus	3.36 (2.5-6.2)	1141 (6.84)
7	Impact on health care practices	change, resident, time, virtual, visit	education, telemedicine, pandemic, health, medical	3.86 (2.5-5.7)	1115 (6.68)
8	Radiology	imaging, tomography, lesion, diagnosis, feature	pneumonia, tomography, computed, disease, ct	3.69 (2.7-5.5)	774 (4.64)
9	Epidemiological modeling	control, spread, measure, public, italy	health, pandemic, model, disease, public	3.48 (2.5-5.2)	1219 (7.31)
10	Clinical manifestations	increase, associate, infection, cardiovascular, injury	disease, acute, syndrome, respiratory, severe	4.99 (3.3-7.8)	1828 (10.96)
11	Protective measures	equipment, high, practice, perform, protective	surgery, cancer, pandemic, management, personal	4.50 (2.6-5.5)	1466 (8.79)
12	Immunology	expression, target, inhibitor, enzyme, viral	ace, angiotensin, protein, molecular, converting	4.56 (3.1-8.1)	694 (4.16)
13	Pregnancy	systematic, datum, include, disease, search	pregnancy, infection, respiratory, transmission, disease	3.52 (2.3-5.1)	819 (4.91)
14	Psychological impact	increase, stress, old, physical, public	health, pandemic, mental, social, anxiety	3.35 (2.4-5.0)	1771 (10.62)

Figure 4 shows the topic distribution over time. In particular, 4 topics (epidemiology, epidemiological modeling, health care response, and radiology) showed a declining curve over time, whereas 2 topics (clinical manifestations and protective measures) showed a clear increase in proportions, and 1 topic (disease transmission) showed a bell-shaped progression.

Figure 4. Proportion of topics in relation to all COVID-19 articles published per week.
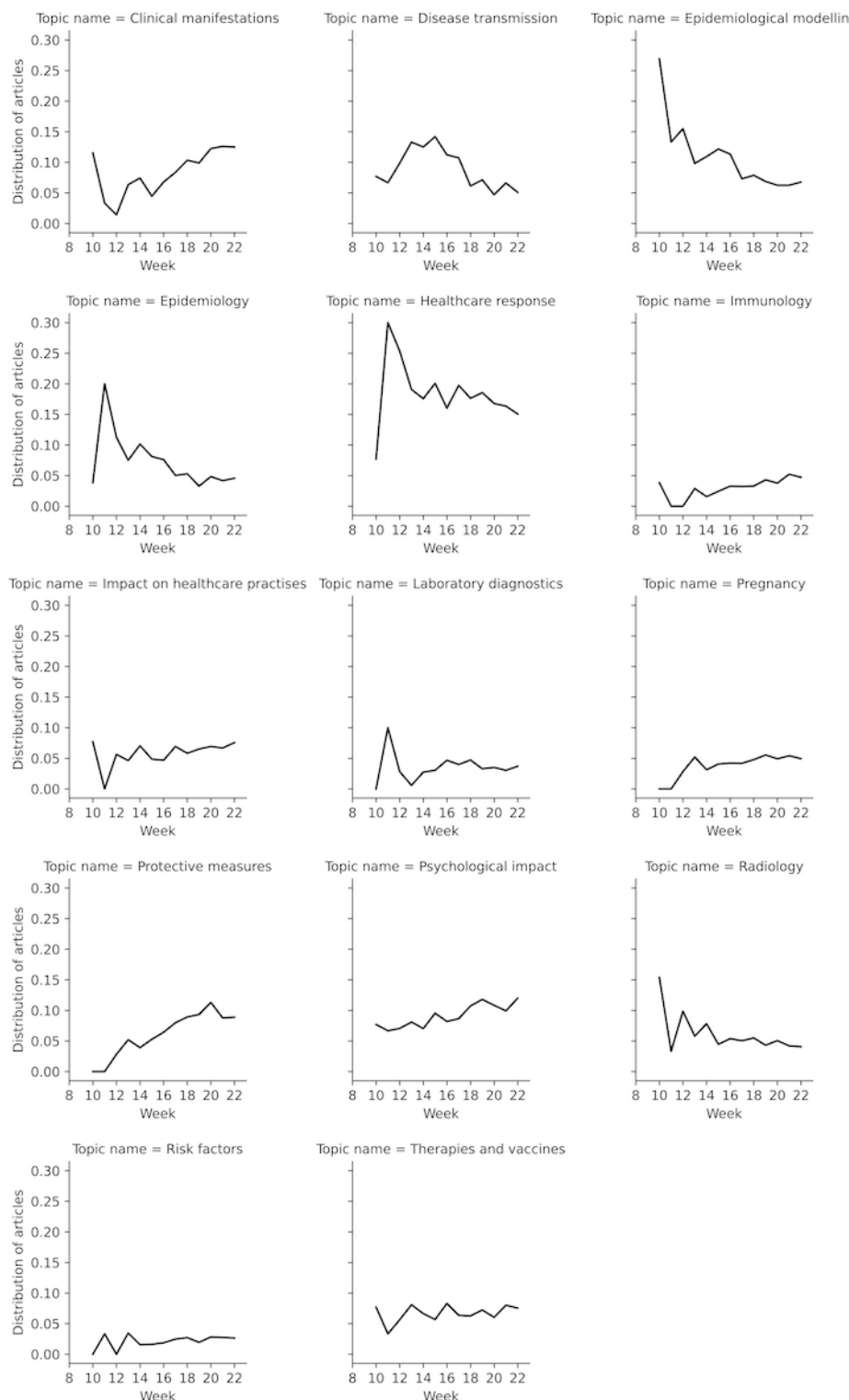
Analysis Based on Journal Impact Factor

In all, the data set comprised 2473 unique journal names. Of these, we found perfect matches for 1129 (45.65%) journal names upon comparison with the Clarivate JIF list. For the remaining 1344 (54.34%) journal names, we used the random forest regression model to compute the JIF. The model showed an RMSE of 5.4 on test data (R^2=47%), indicating that the average prediction erred with 5.4 points (Multimedia Appendix 5).

The median JIF for all articles in the data set was 3.7 (IQR 2.6-5.9). We found a declining trend in median JIF over time (Figure 2). The 3 topics with the highest median JIF were protective measures, immunology, and clinical manifestations (Table 1). There was a low correlation between the median JIF and the number of articles in each topic (Pearson correlation coefficient=−0.14).

Analysis Based on Geographic Origin

Figure 5 shows the geographic origin of the 16,670 identified articles. The top 4 countries of origin were the United States (3223, 19.33%), China (2264, 13.58%), Italy (1591, 9.54%), and the United Kingdom (1055, 6.33%).

Figure 5. Geographic origins of the 16,670 published COVID-19 articles analyzed. Lat: latitude, Long: longitude.
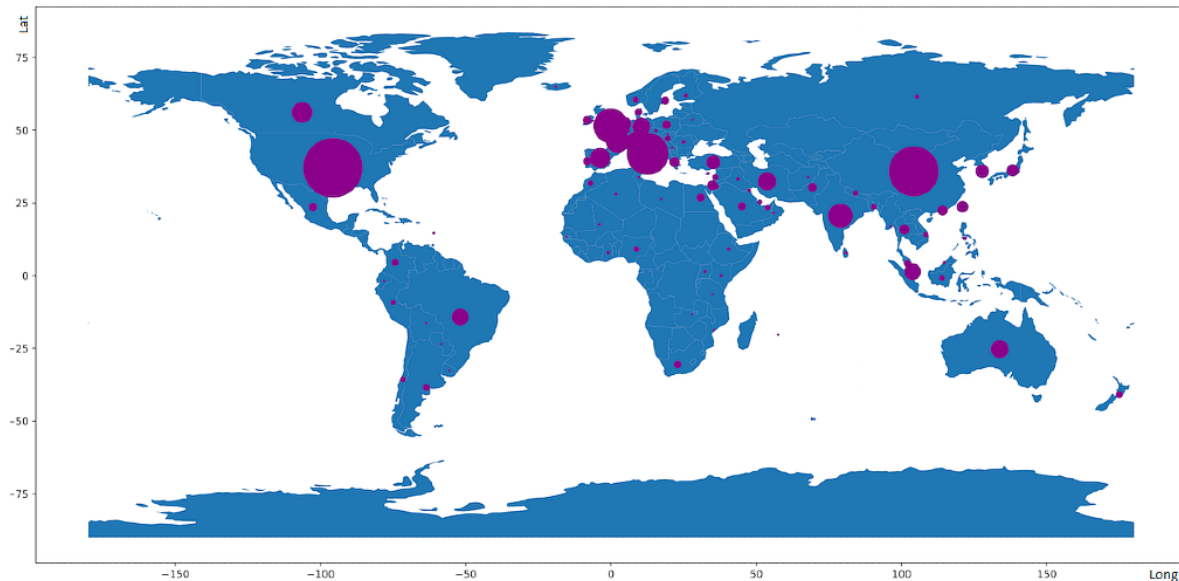
Discussion

We classified the scientific publications on COVID-19 during the early phase of the pandemic into 14 topics. Overall, the most-prevalent topics were health care response, clinical manifestations, and psychological impact. Although the prevalence for some topics, such as health care response, has decreased over time, the prevalence for some other topics, such as clinical manifestations and protective measures, continues to increase. These findings suggest how research priorities have changed over time and, consequently, the topics that researchers consider relevant to study and publish have varied during the course of the pandemic. The 3 topics with the highest median JIF were risk factors, immunology, and clinical manifestations. Clinical manifestation was the only topic that featured in all of the following classifications: (1) the most prevalent topics, (2) the topics with the highest median JIF, and (3) the topics with an increasing prevalence. This finding indicates the importance of research on clinical manifestation during the early course of a pandemic caused by a novel virus. In addition, the results of this study show that the countries responsible for the most scientific outputs were also among the countries worst affected by the COVID-19 pandemic [20].

Moreover, we made a noteworthy finding that, in less than 6 months from the detection of the novel coronavirus, 6 peer-reviewed randomized trials on COVID-19 patients were published. The focus of these trials span from herbal [21] and medical [22-25] treatment options to respiratory rehabilitation [26]. It should be noted that such in-depth analysis of the data is not achieved by the topic model itself but requires some manual control.

When we did a comparative search for scientific articles published during the early phase of the 2009 influenza A (H1N1) pandemic [27], we made a strikingly different observation: The scientific community had a slower reaction to the H1N1 pandemic than to the COVID-19 pandemic, based on the date of publication of the first relevant scientific article, the overall number of publications, and the calculated median JIF. Moreover, the first randomized trial on H1N1 [28] was published (December 17, 2009) 190 days after the outbreak was declared as a pandemic; the corresponding timeframe for the first randomized trial on COVID-19 [23] was 44 days. This comparison, however, may be affected by many fundamental differences between the two pandemics, such as disease novelty and severity.

Previous studies have demonstrated the utility of topic modeling to map online activities [29,30], social media postings [31,32], and media reports on the COVID-19 pandemic [33]. To our knowledge, this is the first study to use topic modeling to assess published research on COVID-19. This study provides an in-depth analysis of a defined short period of time following the emergence of a novel disease. We believe our study findings may serve as an illustration of how the medical research community reacts, what topics are considered to be the most imperative to clarify, how research efforts are distributed geographically, and how they develop over time. Our analysis may also serve to demonstrate how research is being published, for instance, with regard to JIF when a new disease is discovered. Topic modeling enables an assessment of the research evolution, both on short and long term. The method could prove to be suitable for broad fields as well as narrow research questions. Topic modeling may also offer utility for additional in-depth analysis, by further exploring a selected topic to identify and analyze subtopics. Although several systematic reviews on COVID-19 have been published, it should be noted that such reviews do not feature the most recent literature; they are highly time- and resource-consuming; and they generally only focus on a specific aspect of the pandemic [10-12]. Using topic modeling, our study offers a comprehensive

assessment of the available scientific publications on COVID-19.

Our study has some limitations. First, the LDA model does not account for the context of a word, and a certain word may have different meanings depending on the context it is presented in. However, a review of our topics shows that this limitation was not prominent in our data. Therefore, it is unlikely that this limitation of the model negatively affects the interpretation of our topics. Second, there are several quantitative metrics to assess the optimal number of topics, which may conflict with the opinion of a subject matter expert. Any topic model produced, however, should be validated by subject matter experts, because any application of these topics will be done by such experts. Third, an article may appear in several topics in our data set. However, this may be considered a strength, as some overlap is indeed a property of research articles, and the aim of this study was to analyze scientific text in its original form. Fourth, the number of topics could be questioned—fewer topics may be easier for a reader to consume, whereas a larger number of topics could have resulted in a more mathematically optimal solution. However, we believe that a reasonable number of topics needs to be selected to balance mathematical accuracy and utility. Fifth, we limited our analyses to scientific publications. Analyses of grey literature may prove to add important information. Finally, as the COVID-19 pandemic is still ongoing, we expect the research topics to continuously change. To facilitate updated assessments, we have developed a web-based tool using the methods described in this study. Regular updates on the evolution of the COVID-19 evidence base can be found online at the c19research website [34].

Our study findings suggest that the scientific publications during the early phase of the COVID-19 pandemic can be modeled into topics. The evolution of these topics gives insights into current research trends and may aid researchers and policy makers to form a structured view of the existing COVID-19 evidence base and provide further research directions. Furthermore, our findings demonstrate that topic modeling is a rapid and useful method to assess the development of a broad and rapidly evolving research topic, such as COVID-19, and that it could be further utilized during the course of the current as well as future pandemics.

Acknowledgments

Authors' Contributions

MN, AÄ, and OE designed the study. MN and OE collected, cleaned, and interpreted the data. AÄ contributed to the data interpretation and wrote the first draft of the manuscript. MN and OE were major contributors in writing the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Stopwords used in the preprocessing of text data.
[PDF File (Adobe PDF File), 40 KB - jmir_v22i11e21559_app1.pdf]

Multimedia Appendix 2
Top 15 characterizing words for each topic, from the final latent Dirichlet allocation model.
[PDF File (Adobe PDF File), 50 KB - jmir_v22i11e21559_app2.pdf]

Multimedia Appendix 3
Evaluation metrics perplexity and leave-out likelihood for 20 latent Dirichlet allocation models.
[PNG File , 34 KB - jmir_v22i11e21559_app3.png]

Multimedia Appendix 4
Heatmap showing correlation between topic occurrence in the same article.
[PNG File , 112 KB - jmir_v22i11e21559_app4.png]

Multimedia Appendix 5
Histogram showing the distribution of journal impact factor (JIF) for the identified articles.
[PNG File , 30 KB - jmir_v22i11e21559_app5.png]

References

XSL•FO

RenderX

1. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA 2020 Feb 25;323(8):709-710 [FREE Full text] [doi: 10.1001/jama.2020.1097] [Medline: 31999307]

2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]

3. WHO announces COVID-19 outbreak a pandemic. Word Health Organization - Regional Office for Europe. Copenhagen, Denmark URL: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic [accessed 2020-06-01]

4. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]

5. Thorlund K, Dron L, Park J, Hsu G, Forrest JI, Mills EJ. A real-time dashboard of clinical trials for COVID-19. Lancet Digit Health 2020 Jun;2(6):e286-e287. [doi: 10.1016/S2589-7500(20)30086-8] [Medline: 32363333]

6. Global Coronavirus COVID-19 Clinical Trial Tracker. URL: https://www.covid-trials.org/ [accessed 2020-06-01]

7. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]

8. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]

9. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):569-573 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.008] [Medline: 32413821]

10. Li X, Guan B, Su T, Liu W, Chen M, Bin Waleed K, et al. Impact of cardiovascular disease and cardiac injury on in-hospital mortality in patients with COVID-19: a systematic review and meta-analysis. Heart 2020 Aug;106(15):1142-1147 [FREE Full text] [doi: 10.1136/heartjnl-2020-317062] [Medline: 32461330]

11. Parasa S, Desai M, Thoguluva Chandrasekar V, Patel HK, Kennedy KF, Roesch T, et al. Prevalence of gastrointestinal symptoms and fecal viral shedding in patients with coronavirus disease 2019: A systematic review and meta-analysis. JAMA Netw Open 2020 Jun 01;3(6):e2011335 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11335] [Medline: 32525549]

12. Cortegiani A, Ingoglia G, Ippolito M, Giarratano A, Einav S. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. J Crit Care 2020 Jun;57:279-283 [FREE Full text] [doi: 10.1016/j.jcrc.2020.03.005] [Medline: 32173110]

13. Entrez programming utilities help. National Center for Biotechnology Information. Bethesda, MD URL: https://www.ncbi.nlm.nih.gov/books/NBK25501/ [accessed 2020-06-01]

14. Stopwords. PubMed Help [Internet]. Bethesda, MD: National Center for Biotechnology Information URL: https://www.ncbi.nlm.nih.gov/books/NBK3827/table/pubmedhelp.T.stopwords/ [accessed 2020-06-01]

15. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003 Mar 01;3:993-1022 [FREE Full text] [doi: 10.1186/1471-2105-7-250] [Medline: 16681860]

16. Kandula S, Curtis D, Hill B, Zeng-Treitler Q. Use of topic modeling for recommending relevant education material to diabetic patients. AMIA Annu Symp Proc 2011;2011:674-682 [FREE Full text] [Medline: 22195123]

17. Zou C. Analyzing research trends on drug safety using topic modeling. Expert Opin Drug Saf 2018 Apr 06;17(6):629-636. [doi: 10.1080/14740338.2018.1458838] [Medline: 29621918]

18. Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. : Association for Computational Linguistics; 2014 Jun 27 Presented at: Proceedings of the workshop on interactive language learning, visualization, and interfaces; 2014; Baltimore, MD p. 63-70 URL: https://www.aclweb.org/anthology/W14-3110.pdf [doi: 10.3115/v1/W14-3110]

19. 2019 Journal Citation Reports - Full journal list. Web of Science Group. 2019. URL: https://clarivate.com/webofsciencegroup/wp-content/uploads/sites/2/dlm_uploads/2019/08/JCR_Full_Journal_list140619.pdf [accessed 2020-06-01]

20. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University of Medicine - Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-06-01]

21. Hu K, Guan W, Bi Y, Zhang W, Li L, Zhang B, et al. Efficacy and safety of Lianhuaqingwen capsules, a repurposed Chinese herb, in patients with coronavirus disease 2019: A multicenter, prospective, randomized controlled trial. Phytomedicine 2020 May 16:153242. [doi: 10.1016/j.phymed.2020.153242] [Medline: 32425361]

22. Cao Y, Wei J, Zou L, Jiang T, Wang G, Chen L, et al. Ruxolitinib in treatment of severe coronavirus disease 2019 (COVID-19): A multicenter, single-blind, randomized controlled trial. J Allergy Clin Immunol 2020 Jul;146(1):137-146.e3 [FREE Full text] [doi: 10.1016/j.jaci.2020.05.019] [Medline: 32470486]

23. Borba MGS, Val FFA, Sampaio VS, Alexandre MAA, Melo GC, Brito M, CloroCovid-19 Team. Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: a randomized clinical trial. JAMA Netw Open 2020 Apr 24;3(4):e208857 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8857] [Medline: 32330277]

24.   Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. The Lancet 2020 May 16;395(10236):1569-1578 [FREE Full text] [doi: 10.1016/S0140-6736(20)31022-9] [Medline: 32423584]

25.   Hung IF, Lung K, Tso EY, Liu R, Chung TW, Chu M, et al. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. The Lancet 2020 May 30;395(10238):1695-1704 [FREE Full text] [doi: 10.1016/S0140-6736(20)31042-4] [Medline: 32401715]

26.   Liu K, Zhang W, Yang Y, Zhang J, Li Y, Chen Y. Respiratory rehabilitation in elderly patients with COVID-19: a randomized controlled study. Complement Ther Clin Pract 2020 May;39:101166 [FREE Full text] [doi: 10.1016/j.ctcp.2020.101166] [Medline: 32379637]

27.   Pandemic influenza A (H1N1) Donor Report March 1, 2011. World Health Organization. Geneva, Switzerland; 2011. URL: https://www.who.int/csr/resources/publications/swineflu/h1n1_donor_032011.pdf [accessed 2020-06-01]

28.   Greenberg ME, Lai MH, Hartel GF, Wichems CH, Gittleson C, Bennet J, et al. Response to a monovalent 2009 influenza A (H1N1) vaccine. N Engl J Med 2009 Dec 17;361(25):2405-2413. [doi: 10.1056/NEJMoa0907413] [Medline: 19745216]

29.   Jo W, Lee J, Park J, Kim Y. Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis. J Med Internet Res 2020 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]

30.   Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public priorities and concerns regarding COVID-19 in an online discussion forum: longitudinal topic modeling. J Gen Intern Med 2020 Jul;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]

31.   Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]

32.   Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

33.   Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]

34.   The number of publications, journal impact factor, and research topics since January 2020. The evolution of COVID-19 research. URL: http://www.c19research.org [accessed 2020-06-01]

Abbreviations

JIF: journal impact factor
LDA: latent Dirichlet allocation
PCA: principal component analysis
RMSE: root mean square error

Original Paper

Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation

Pan Pan[1], MD; Yichao Li[2], BSc; Yongjiu Xiao[3], MD; Bingchao Han[4], MD; Longxiang Su[5], MD; Mingliang Su[2], BSc; Yansheng Li[2], BSc; Siqi Zhang[2], BSc; Dapeng Jiang[2], BSc; Xia Chen[2], BSc; Fuquan Zhou[2], BSc; Ling Ma[3], BSc; Pengtao Bao[6], MD; Lixin Xie[6], MD

[1]Chinese PLA General Hospital, Medical School Of Chinese PLA, College of Pulmonary and Critical Care Medicine, Beijing, China
[2]DHC Mediway Technology Co Ltd, Beijing, China
[3]The 940th Hospital of Jiont Logistics Support Force of Chinese People's Liberation Army, Lanzhou, China
[4]The 980th Hospital of Jiont Logistics Support Force of Chinese People's Liberation Army, Shijiazhuang, China
[5]Peking Union Medical College Hospital, Beijing, China
[6]College of Pulmonary and Critical Care Medicine, Chinese PLA General Hospital, Beijing, China

Corresponding Author:
Lixin Xie, MD
College of Pulmonary and Critical Care Medicine
Chinese PLA General Hospital
No.28 Fuxing Road, Haidian District, Beijing
Beijing
China
Phone: 86 55499128
Email: xielx301@126.com

Abstract

Background: Patients with COVID-19 in the intensive care unit (ICU) have a high mortality rate, and methods to assess patients' prognosis early and administer precise treatment are of great significance.

Objective: The aim of this study was to use machine learning to construct a model for the analysis of risk factors and prediction of mortality among ICU patients with COVID-19.

Methods: In this study, 123 patients with COVID-19 in the ICU of Vulcan Hill Hospital were retrospectively selected from the database, and the data were randomly divided into a training data set (n=98) and test data set (n=25) with a 4:1 ratio. Significance tests, correlation analysis, and factor analysis were used to screen 100 potential risk factors individually. Conventional logistic regression methods and four machine learning algorithms were used to construct the risk prediction model for the prognosis of patients with COVID-19 in the ICU. The performance of these machine learning models was measured by the area under the receiver operating characteristic curve (AUC). Interpretation and evaluation of the risk prediction model were performed using calibration curves, SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), etc, to ensure its stability and reliability. The outcome was based on the ICU deaths recorded from the database.

Results: Layer-by-layer screening of 100 potential risk factors finally revealed 8 important risk factors that were included in the risk prediction model: lymphocyte percentage, prothrombin time, lactate dehydrogenase, total bilirubin, eosinophil percentage, creatinine, neutrophil percentage, and albumin level. Finally, an eXtreme Gradient Boosting (XGBoost) model established with the 8 important risk factors showed the best recognition ability in the training set of 5-fold cross validation (AUC=0.86) and the verification queue (AUC=0.92). The calibration curve showed that the risk predicted by the model was in good agreement with the actual risk. In addition, using the SHAP and LIME algorithms, feature interpretation and sample prediction interpretation algorithms of the XGBoost black box model were implemented. Additionally, the model was translated into a web-based risk calculator that is freely available for public usage.

Conclusions: The 8-factor XGBoost model predicts risk of death in ICU patients with COVID-19 well; it initially demonstrates stability and can be used effectively to predict COVID-19 prognosis in ICU patients.

Introduction

COVID-19 is a new and severe infectious disease that has spread to 34 provinces and cities in China and over 30 countries worldwide [1,2]. After the entire nation of China fought against COVID-19, by early May 2020, the numbers of patients with COVID-19 had greatly decreased in almost all provinces and cities in China. However, in mid-June, a new outbreak of COVID-19 cases occurred in Beijing, the capital of China. Government efforts have now brought the overall spread of COVID-19 under control. It is clear that COVID-19 is an infectious disease that requires ongoing attention from the medical community, governments, and the public to prevent future outbreaks. As of the end of June 2020, more than 10,000,000 COVID-19 cases had been recorded worldwide. Therefore, an evaluation and early warning system for COVID-19 prognosis is urgently needed, especially for critically ill patients.

COVID-19 cases are classified as mild, moderate, severe, or critical [3]. At present, most studies of COVID-19 have focused on risk factor analysis and mortality prediction for mild and moderate cases, which comprise a large proportion of patients with COVID-19 [4-8]. However, 14% to 20% of cases are severe or even critical [1,9], and the mortality rate of these patients is as high as 50% [10]. Few studies have reported risk factor prediction and mortality analysis for severe and critical patients with COVID-19. COVID-19 predictive models are rapidly entering the academic literature. These include predictive models that are mainly used to identify high-risk groups in the general population [11-13], diagnostic models that are used to detect COVID-19 [14-16], and models used to predict mortality, serious disease progression, etc [17-20]. The most common predictors of the diagnosis and prognosis of COVID-19 are age, body temperature, lymphocyte count, and lung imaging characteristics. The estimated C indices of these predictive models are between 0.65 and 0.99. Although the estimated C indices of some models appear to be ideal, all the models are rated as being at high risk of bias, mainly because of the high risk of model overfitting. Moreover, many of the report descriptions are vague. Most reports do not include a description of the study population or the intended use of the model, and very few evaluations of the calibration of model predictions were made [21].

The theoretical core of machine learning analysis is the data mining algorithm. Various data mining algorithms based on different data types and formats can more scientifically represent the characteristics of the data and can better penetrate the data trends and recognized values [22]. On this basis, one of the most important application areas is predictive analysis, which involves identifying features (in machine learning, "features" refers to individual characteristics of the data) from mechanical learning, establishing models through science, and then running new data through the models to predict future data [23]. In this study, we clarify that the established model is used to predict the prognosis of patients with COVID-19 in the intensive care unit (ICU).

The model must be continuously optimized and evaluated. In terms of evaluation, we checked the accuracy and calibration of the model. Moreover, to improve the interpretability of the black box model, we also used SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) to explain the prediction model; therefore, the prediction model not only predicts prognostic outcomes but also gives a reasonable explanation for the prediction, which can greatly enhance users' trust of the model.

Methods

Study Design and Data Source

Vulcan Hill Hospital, located in Wuhan, Hubei Province, is a special hospital that was built by the Chinese government to treat patients with COVID-19. Construction on the hospital started on January 24, 2020, and was completed on February 1; the hospital entered use on February 2, and it officially closed on April 15. During this period, a total of 3063 patients with laboratory-confirmed hospitalized cases of COVID-19 were admitted. For this study, data of 3063 patients with COVID-19 admitted to Vulcan Hill Hospital were extracted from hospital medical records and screened for eligibility. The study extracted 100 relevant variables, such as baseline patient information, clinical diagnosis, vital signs, laboratory test results, medical advice, and nursing care, as candidate variables for predictors [24]. We established a study cohort of 123 critically ill patients admitted to the ICU, and 2940 patients who did not enter the ICU were excluded. Considering the problem of predictors and the timing of the outcome measurement, we used the time the patient entered the ICU to calculate the first test value of all candidate predictors upon entering the ICU. The output of our study is the prognostic outcome of these critically ill patients, and this outcome is based on the ICU death record in the electronic medical record. After further checking the admission records, we included data of 123 critically ill patients admitted to the ICU, including 65 (52.8%) who survived and 58 (47.2%) who died. We randomly used 80% of these data as the training set, and the remaining 20% were used as the validation set. For the training set data, we completed the statistics and proper processing of missing values, the identification and processing of noise data, and the standardization of all predictive variables. The validation set was processed in exactly the same way as the training set. Then, we researched and analyzed the feature selection, model training, model evaluation, and model interpretation.

It is worth noting that this study adhered to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement for reporting, and completion of the model construction and verification was guided by PROBAST (Prediction model Risk Of Bias ASsessment Tool) [25,26]. This study was approved by the ethics committee of Vulcan Hill Hospital; the requirement for informed consent was waived.

Predictor Variables and Data Preprocessing

In this study, a total of 100 candidate predictive features were collected, and the test results were the first measured value after the patient entered the ICU. Among these features, 5 features with a missing ratio greater than 30% were excluded, and the remaining features were filled with missing values using appropriate methods. Because missing data may lead to loss of useful information and even create instability of the model realization, it is more difficult to analyze the model results with missing data; therefore, we carried out a cautious missing value interpolation strategy. We used the Iterative Imputer tool developed by scikit-learn to perform multiple imputations for missing values. The Iterative Imputer uses an algorithm to model each missing value feature as a function of other features. It uses the predicted value of the function as an estimate. In each step, one feature is selected as the output y, and all other features are selected as the input X. Then, a regressor is trained on X and y to predict the missing value of y. The area under the receiver operating characteristic curve (AUC) values corresponding to each padding method were found to be basically the same. We also applied the K Neighbors Regressor, Decision Tree Regressor, Bayesian Ridge, and Extra Trees Regressor regression algorithms as predictors to complete missing value filling. Moreover, we attempted mean filling and median filling and fed the above six imputation results into the traditional logistic regression model to calculate and compare the areas under the receiving operator characteristic curve (AUC) of their respective prediction results. The results in Figure 1 show that the best filling method is multiple imputation, with Decision Tree Regressor as the regression method; therefore, this method was finally used to fill in the missing data for the continuous features. The missing data are provided in Multimedia Appendix 1. In addition, we drew box plots for continuous features and used IQR criteria to filter and replace outliers. Finally, to obtain more reliable prediction performance, the continuous data were standardized by the z score standardization method.

Figure 1. (A) Comparison of the AUC values obtained after using logistic regression as a carrier and filling in discrete features with multiple missing filling methods. (B) Comparison of the AUC values obtained after continuous feature filling by multiple missing filling methods using logistic regression as a carrier. AUC: area under the receiver operating characteristic curve.
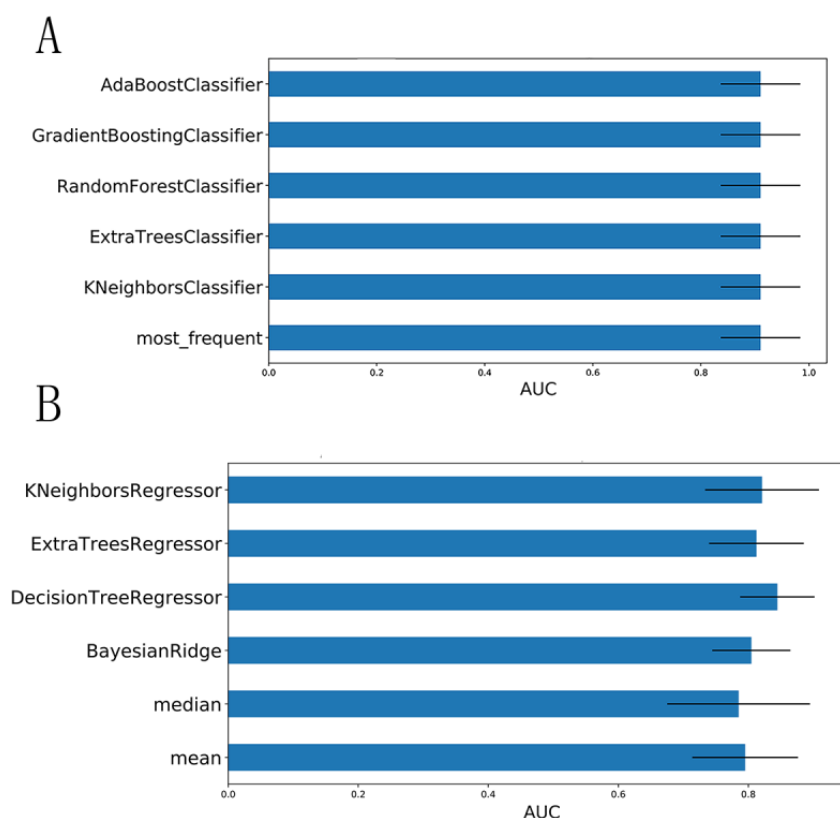
Feature Selection and Statistical Analysis

The categorical variables of the queue data were expressed as n (%). Continuous variables that satisfy normal distribution were expressed as mean (SD); otherwise, medians and quartiles were used. All characteristics were evaluated statistically, and two-sided differences in P values <.05 were considered statistically significant. Differences between categorical variables were compared using the chi-square test or Fisher exact test as needed. The independent sample t test was used to compare continuous variables that satisfy normal distribution, while the Wilcoxon test was used for nonnormally distributed continuous variables. Statistically significant features were selected for further correlation analysis. For redundant features with strong correlation, factor analysis was used to confirm the collinearity of the variables, classify the collinearity as a latent factor, and then calculate the eigenvalues, visualize the gravel map of the eigenvalues, and select the feature root. Values >1

and the first few principal components where the slope decreases were used as principal component factors to eliminate redundant features and to find more efficient, concise, and precise feature combinations, thereby improving the generalization and practical capabilities of the model, as described previously [27]. Note that factor analysis requires the data to be suitable according to the Bartlett sphere test and Kaiser-Meyer-Olkin test.

Derivation and Validation of the Models

A conventional logistic regression method and four popular machine learning classification algorithms, including adaptive boosting (AdaBoost), gradient boosting decision tree (GBDT), eXtreme Gradient Boosting (XGBoost), and CatBoost, were applied in the present study to model the data. The model built by the algorithm uses constant parameter optimization and model evaluation to compare the fitting effects of each model and to select the best model as the risk prediction model. Model optimization is a method that combines grid search and five-fold cross-validation to visualize the AUC values of the model and the standard deviation with the parameters and selects the parameter values corresponding to the best AUC value as the model parameters [28].

A good model explanation must be presented for the black box model. This study is based on the SHAP algorithm, which calculates the marginal contribution of a feature when it is added to the model and then considers whether the factor is different in all factor sequences [29]. The marginal contribution fully explains the influence of all factors included in the model for model prediction and distinguishes the attributes of the factors (risk factors and protective factors).

Finally, validation queue data were used to evaluate the prediction performance of the model and calculate the AUC, threshold, Youden index, 95% CI, SD, and P value of the AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) of the model for the test set. A calibration curve was drawn, the calibration degree of the model was measured, and the degree of consistency between the predicted risk and the actual risk of the model was evaluated. A good calibration degree indicates that the predicted value of the model is closer to the actual probability of the outcome, from the interpretation of the model to the prediction of random samples, as demonstrated previously [30]. The model, code, and parameters are provided in Multimedia Appendix 2.

Interpretation of the Model for Prediction of Random Samples

Advanced machine learning models are usually "black boxes." When the internal operations of a model are unknown, users do not trust the reliability of the model for making predictions. Although it is known that the accuracy of cross-validation of these models is very high, correlation is still sometimes found between the verification data and the training data due to improper methods, especially when there are few samples. Therefore, cross-validation is no longer the only indicator for evaluating trust. If the rationale by which a model predicts a single sample can be intuitively perceived, users can better trust or distrust single sample prediction. The LIME algorithm was implemented with this concern in mind. This linear model is used to locally approximate a black box model by giving weights to the disturbance input; thus, the observation model gives a basis for interpretation of the sample prediction results [31]. In the present study, we randomly sampled the test set and used the LIME algorithm to fit the predictive behavior of the model to the sample to verify the rationality of the basis of the model for predicting results.

Results

Study Population and Baseline Characteristics

Data from 3063 patients with COVID-19 treated at Vulcan Hill Hospital from February 2 to April 15, 2020, were analyzed retrospectively. A total of 69/3063 deaths occurred (2.3%). The final analytic sample included 123 critically ill patients admitted to the ICU, including 85 critically ill patients (69.1%), 65 surviving patients (52.8%), and 58 patients who died (47.2%). The outcome variable was determined as the prognostic outcome of critically ill patients, and the outcome was based on the death record in the electronic medical record. 100 related variables were established as candidate variables for predictors. Figure 2 shows a flowchart of the overall process of data and feature screening. Figure 3 and Multimedia Appendix 3 list the results comparing all potential risk factors in the study cohort. Overall, the mean age of the patients in the cohort was 69.8 years (SD 11.1), and 79/123 patients (64.2%) were male. Data analysis revealed significant differences between patients in 7 discrete factors, namely ventilator use, critical illness, vasoactive drugs, carbapenem use, antibiotic resistance, anti–gram-positive cocci, and hemodiafiltration; significant differences were found in 46 continuous factors.

Figure 2. Flowchart of the data and feature selection.
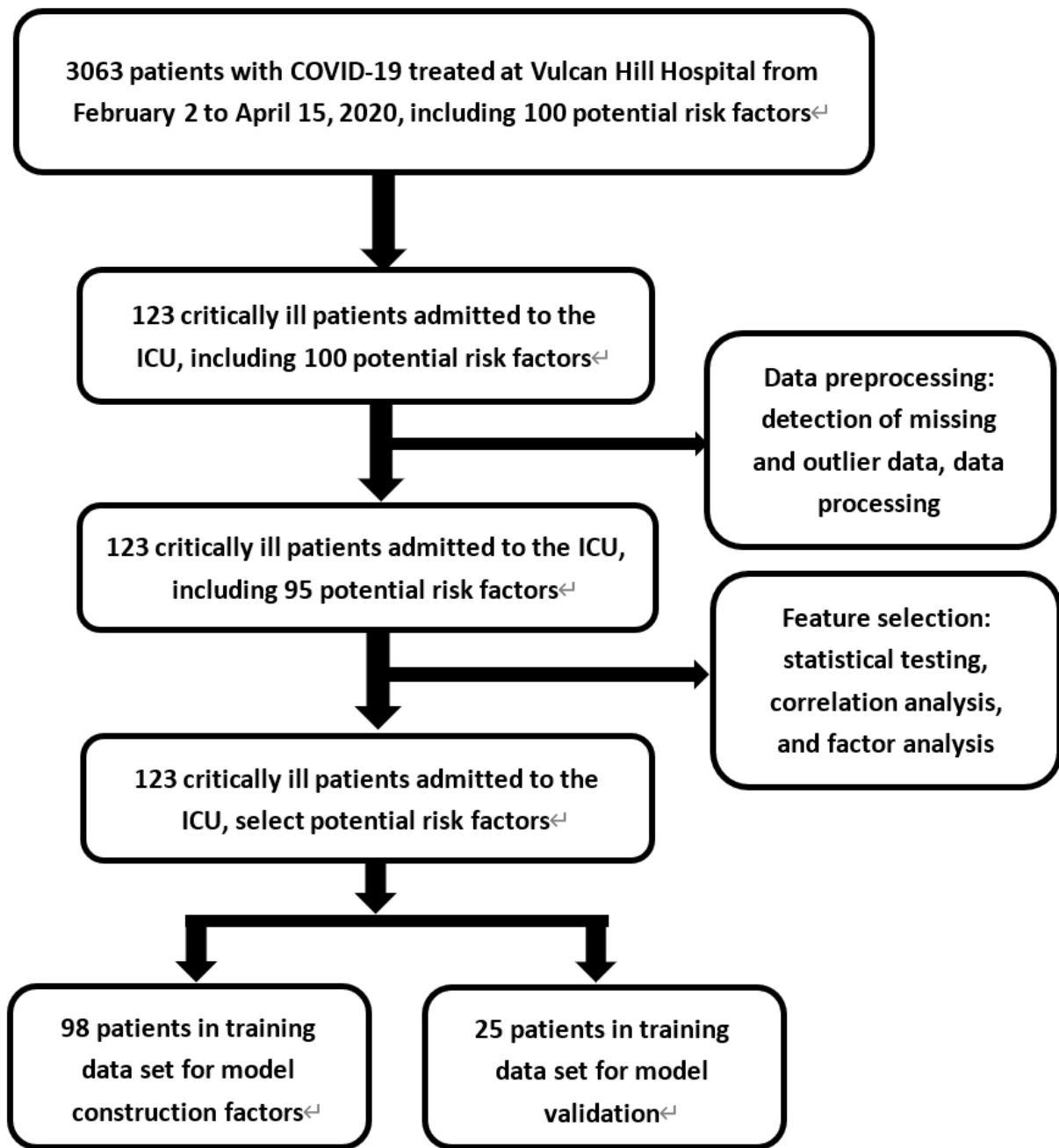
Figure 3. Statistical analysis of the screening results of all potential risk factors in the study cohort. The figure shows the screening of candidate variables in the five dimensions of patient baseline information, clinical diagnosis and vital signs, laboratory tests, medical advice, and nursing care.
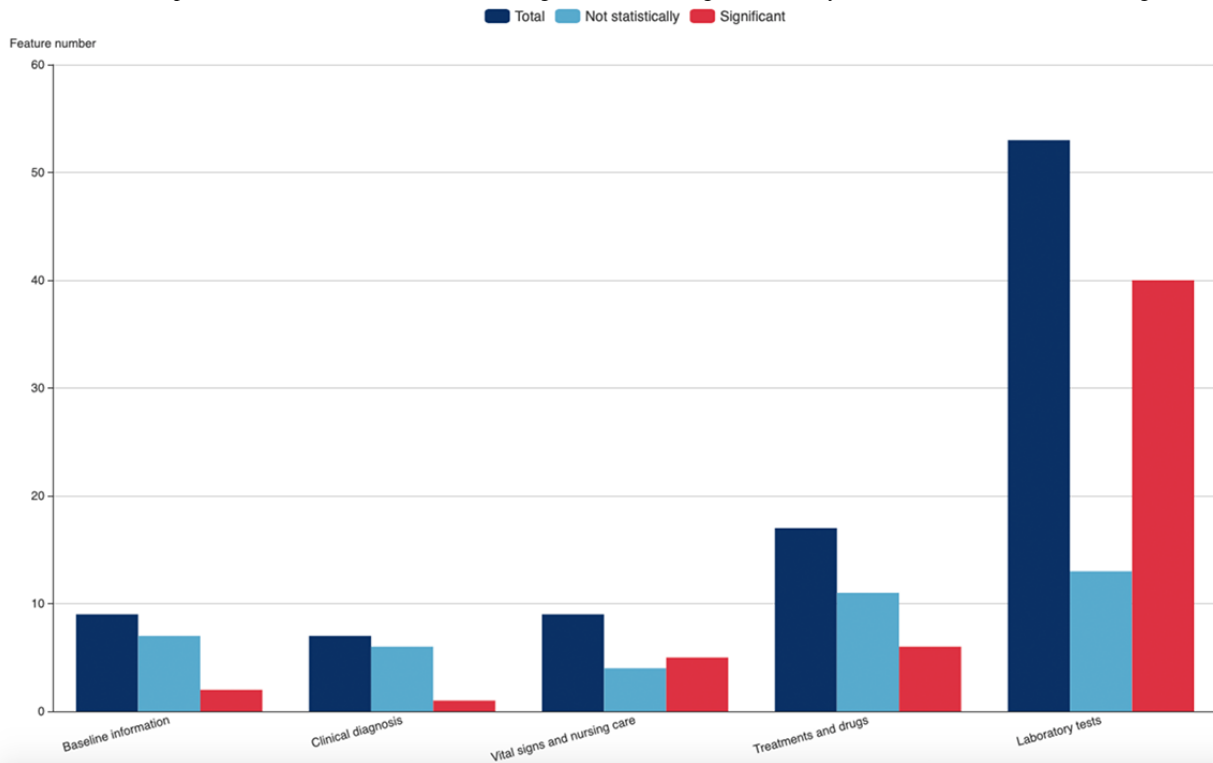
Predictor Selection

By selecting the abovementioned statistically significant factors for correlation analysis, the correlation coefficient matrix heat map (Multimedia Appendix 4) of the features shows that the top five features that were negatively correlated with the outcomes are prothrombin time percentage activity, blood oxygen saturation, lymphocyte percentage (LYM%), albumin level (ALB), and percentage of basophils (BASO%); the top five characteristics that were positively correlated with outcomes are lactate dehydrogenase, alpha-hydroxybutyrate dehydrogenase, C-reactive protein, neutrophil percentage (NEUT%), and original thrombin time. In addition, strong correlations were found between many features. For example, the correlation coefficient between the prothrombin time (PT)

and international standardized ratio reached 0.999; therefore, it was necessary to reduce redundant features.

Factor analysis and visualization of the characteristic root gravel map and load matrix (Figure 4 and Multimedia Appendix 5) revealed that the eight principal component factors were the most predictive; for example, the correlation between the characteristic prothrombin time and the second main factor reached 0.97. Considering the convenience and practicability of using the prediction model, clinical experience and actual comparisons were combined to finally select eight features to represent the eight principal component factors, namely LYM%, PT, lactate dehydrogenase (LDH), total bilirubin (T-Bil) , eosinophil percentage (EOS%), creatinine (Cr), NEUT%, and ALB. The Kaiser-Meyer-Olkin test gave a value of 0.5714 and Bartlett's test of sphericity showed a significance level of $P<.001$, indicating that the factor analysis is effective.

Figure 4. Distribution diagram of the correlations between the feature value and the number of features; when the feature value is >1 and the slope change becomes slow, the number of features is 8.
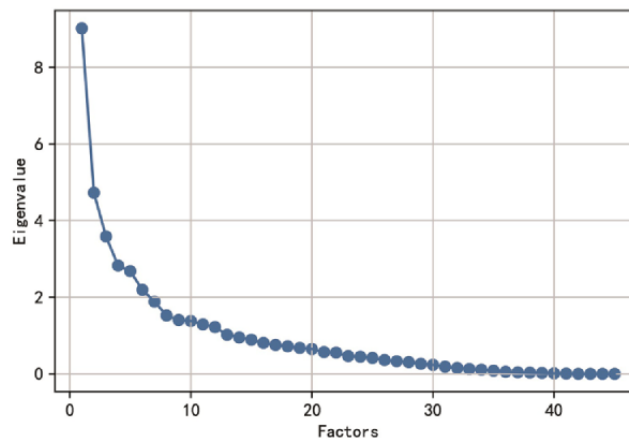
Machine Learning Algorithm Comparison and Best Model

Comparing the AUCs of the logistic regression and four machine learning algorithms for 5-fold cross-validation on the training set (Figure 5), it can be found that the AUC values of each algorithm are similar; however, the AUC value of the XGBoost algorithm is higher. The XGBoost algorithm reflects a good learning curve on the training set, effectively preventing overfitting. In terms of prediction performance, the results of the logistic regression and four machine learning algorithms on the test set show AUCs of 0.92 for XGBoost, 0.9133 for CatBoost, 0.9133 for AdaBoost, 0.85 for GBDT, and 0.84 for LR; the best prediction performance was observed with XGBoost. In addition, the AUC, threshold, Youden index, 95% CI, SD, and P value of the AUC, accuracy, sensitivity, specificity, PPV, NPV, PLR, and NLR values of each model in the test data set are listed in Table 1. In summary, the results of the test data set show that the XGBoost model demonstrates the best performance based on eight salient features. The Youden index value of this model is 0.6667.

Figure 5. ROC curves showing the fitting performance (A) and prediction performance (B) of the LR, CatBoost, GBDT, XGBoost, and AdaBoost prediction models based on the eight important features in the training data set and the test data set. (C) The learning curve on the training set showing the learning process of the XGBoost model. The red line is the fitting effect of the model on the overall training set, and the green line is the fitting effect of the model on the training set with 5-fold cross-validation. The two curves finally merge near 0.85, indicating that the model is well fitted for training. AdaBoost: adaptive boosting; GBDT: gradient boosting decision tree; LR: logistic regression; ROC: receiver operating characteristic; XGBoost: eXtreme Gradient Boosting.
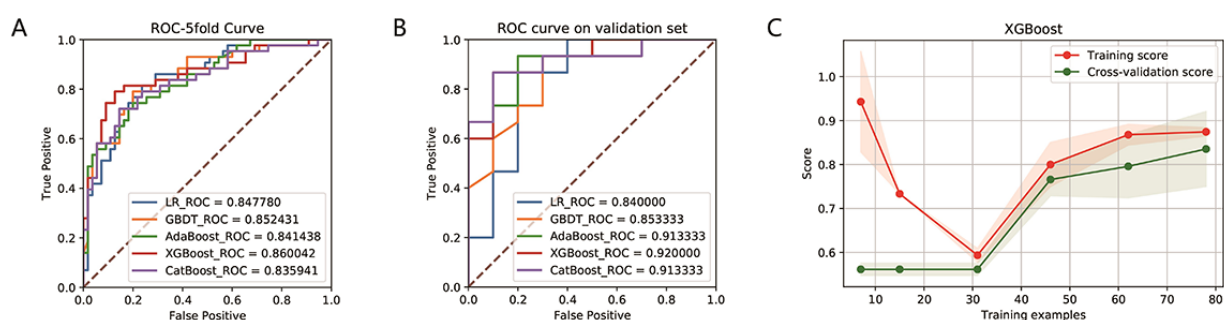
Table 1. Summary of prediction results of multiple models on the test set.

Value	Models				
	Logistic regression	AdaBoost[a]	GBDT[b]	XGBoost[c]	CatBoost
AUC[d]	0.84	0.9133	0.85	0.92	0.9133
Threshold	0.3962	0.4283	0.4583	0.4478	0.5063
Youden index	0.6667	0.7333	0.6333	0.7667	0.7667
95% CI of the AUC	0.6556-1.0	0.8024-1.0	0.6997-1.0	0.8142-1.0	0.7997-1.0
SD of the AUC	0.094	0.0566	0.0784	0.054	0.058
P value of the AUC	.003	<.001	.002	<.001	<.001
Accuracy	0.76	0.76	0.76	0.84	0.84
Specificity	0.8	0.9	0.8	0.9	0.8
Sensitivity	0.7333	0.6667	0.7333	0.8	0.8667
Positive predictive value	0.8462	0.9091	0.8462	0.9231	0.8667
Negative predictive value	0.6667	0.6429	0.6667	0.75	0.8
Positive likelihood ratio	3.6667	6.6667	3.6667	8	4.3333
Negative likelihood ratio	0.3333	0.3704	0.3333	0.2222	0.1667

[a]AdaBoost: adaptive boosting.

[b]GBDT: gradient boosting decision tree.

[c]XGBoost: eXtreme Gradient Boosting.

[d]AUC: area under the receiver operating characteristic curve.

Model Validation and Predictor Parameters

The prediction behavior of XGBoost in the test set was visualized. The calibration curve (Figure 6) shows that the predicted risk of the XGBoost model is in good agreement with the actual risk. The predicted value of the model is close to the actual probability of the outcome. The details of the optimal

XSL•FO

RenderX

model parameters constructed by the XGBoost algorithm can   be viewed in Multimedia Appendix 2.

Figure 6. Calibration curve reflecting the degree of consistency between the predicted risk and the actual risk of the XGBoost model. The predicted curve of the model fits well with the diagonal, indicating that the predicted value of the model is basically close to the actual probability of the outcome.
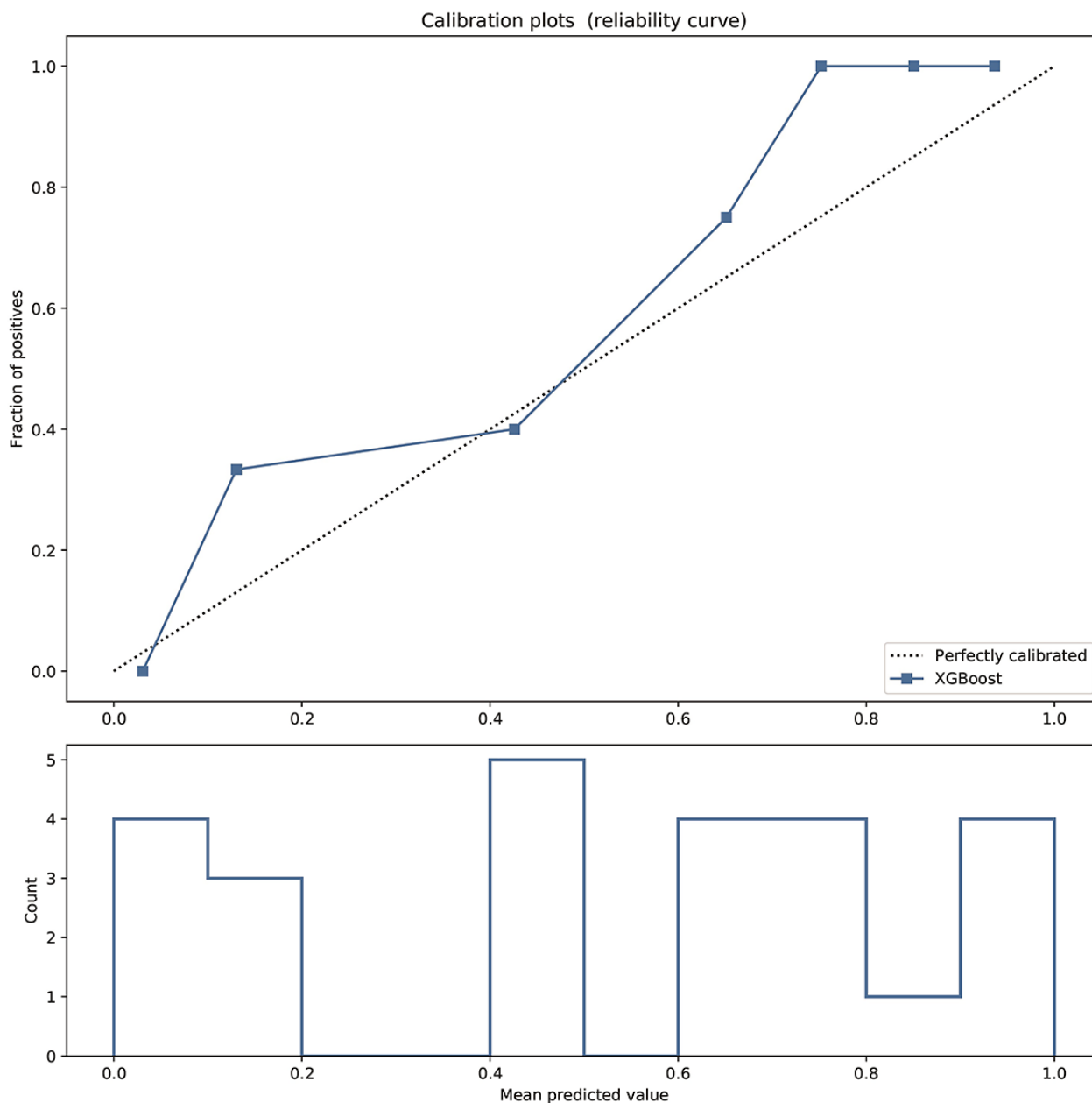
Interpretation and Evaluation of the Machine Learning Model

Based on the SHAP algorithm, the feature ranking interpretation of the XGBoost model (Figure 7) shows that LDH, PT, Cr, LYM%, NEUT%, EOS%, T-Bil, and ALB were the characteristics of the XGBoost model with the greatest impact in predicting outcomes. Overall, the characteristics of LDH, PT, Cr, T-Bil, and NEUT% correlated positively with the outcomes and are risk factors; meanwhile, LYM%, EOS%, and ALB correlated negatively with the outcomes and are protective factors.

Figure 7. The XGBoost model based on the SHAP algorithm. (A) The attributes of the features in the black box model. Each line represents a feature, and the abscissa is the SHA*P* value, which represents the degree of influence on the outcome. Each dot represents a sample. The redder the color, the greater the value of the feature, and the bluer the color, the lower the value. (B) Ranking of feature importance indicated by SHAP. ALB: albumin level; Cr: creatinine; EOS%: eosinophil percentage; LDH: lactate dehydrogenase; LYM%: lymphocyte percentage; NEUT%: neutrophil percentage; PT: prothrombin time; SHAP: SHapley Additive exPlanations; T-Bil: total bilirubin.
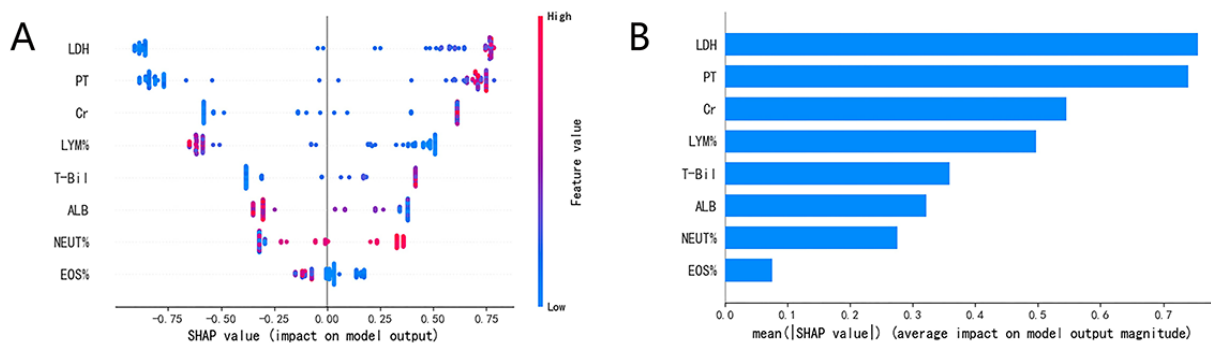
Interpretation of sample prediction results requires random drawing of samples to make model predictions and observe the model through the LIME algorithm. The four prediction scenarios are shown in Figure 8A. In addition, the prediction results of the XGBoost model on all samples on the training set and the test set were counted, and the distribution of the four cases of the aggregated prediction results was visualized. Figure 8B shows that the XGBoost model has inappropriate prediction behavior in that its judgment of the false positive prediction results found through LIME is inaccurate; however, this situation is very rare, which indicates that the performance of the XGBoost prediction model is stable and reliable and that the interpretation of random sample prediction is basically reasonable. This is sufficient to confirm the practicability of the XGBoost model and will help increase physicians' trust in the prediction model and help them make good auxiliary decisions.

Figure 8. Interpretation of sample prediction results by randomly drawing samples to make model predictions and observing the model through the Local Interpretable Model-Agnostic Explanations (LIME) algorithm. (A) The four different prediction behaviors of the model (true negative, true positive, false negative, and false positive); (B) the ratios of the four prediction behaviors of the model on the training set and the test set. A:LB: albumin level; CR: creatinine: EOS%: eosinophil percentage; LDH: lactate dehydrogenase; LYM%: lymphocyte percentage; NEUT%: neutrophil percentage; PT: prothrombin time; T-Bil: total bilirubin.

Additionally, comparison of the results of our machine learning model and the Acute Physiologic Assessment and Chronic Health Evaluation II (APACHE II), Sequential Organ Failure Assessment (SOFA), Multiple Organ Disfunction Score (MODS), and Pneumonia Severity Index (PSI) scores indicated that the AUC of the XGBoost model was higher than those of the other four scores (Figure 9).

Figure 9. Comparison between the risk prediction model of the present study and the ROC curves of various critical scores. APACHE: Acute Physiologic Assessment and Chronic Health Evaluation; MODS: Multiple Organ Disfunction Score; PSI: Pneumonia Severity Index; ROC: receiver operating characteristic; SOFA: Sequential Organ Failure Assessment; XGBoost: eXtreme Gradient Boosting.
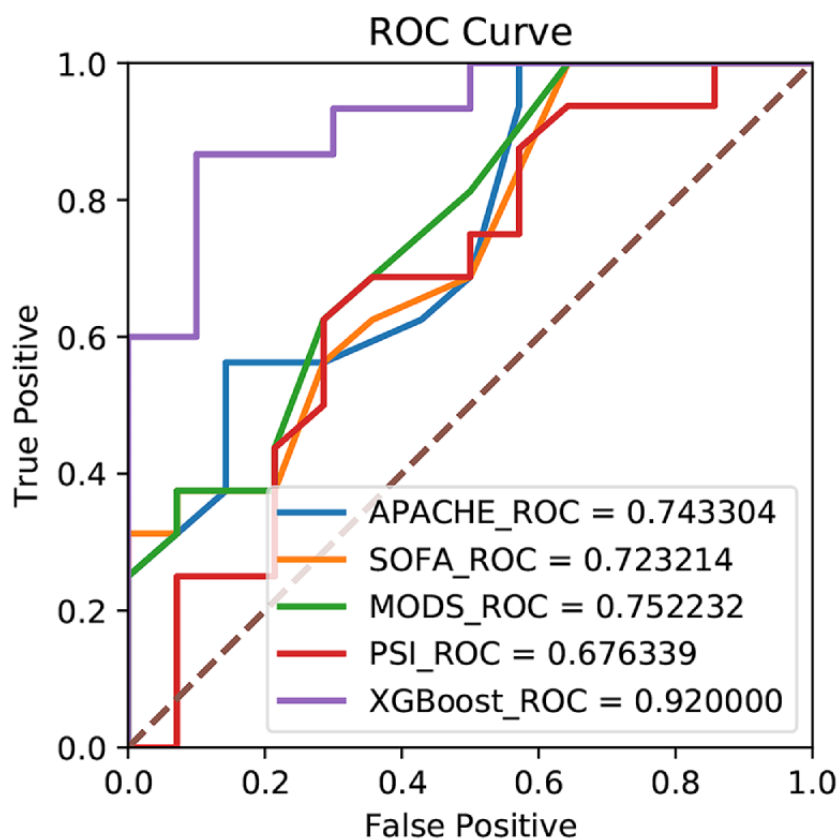
Discussion

Principal Findings

The results of the present study show that the XGBoost method is a more reliable and more accurate method for predicting outcomes for critically ill patients with COVID-19 in the ICU than conventional logistic regression and scoring. Especially, the eigenvalues were reduced using the XGBoost model from 100 parameters to 8. Correlation analysis and characteristic analysis showed that the LDH, PT, Cr, T-Bil, LYM%, ALB, NEUT%, and EOS% indicators had strong correlations with the prognosis of severe and critical patients with COVID-19 in the ICU. Physicians should be wary of poor prognosis when encountering such patients. After full verification by SHAP, LIME, etc, the model was found to be accurate and stable. A web-based calculator based on the risk model is available on the internet [32].

According to the XGBoost algorithm model used in our study, LDH, PT, Cr, T-Bil, and NEUT% correlated positively with patients' outcomes, indicating that these values are risk factors; meanwhile, LYM%, ALB, and EOS% correlated negatively with patients' outcomes, indicating that these values are protective factors. In addition to identifying risk and protective factors, the results suggested that the XGBoost algorithm model achieves a good prediction effect, with an AUC of 0.92, sensitivity of 0.8, and specificity of 0.9. Based on the same data, logistic regression analysis showed an AUC of 0.84, with a sensitivity of 0.7 and specificity of 0.83. These results indicate that the predictive effect of machine learning is more accurate and sensitive than that of regression analysis. In contrast to our study, other researchers tended to apply Cox regression and logistic regression to analyze risk factors. Wang et al [19] found that the risk factors for in-hospital mortality from COVID-19 were lymphopenia and LDH, as analyzed by multivariable Cox proportional hazard regression models. Chen and colleagues [20] studied 1859 patients with confirmed COVID-19 from seven centers in Wuhan, China, of whom 1651 recovered and 208 died. Multivariable Cox regression analyses indicated that increased hazards of in-hospital death were associated with one indicator, log 10 serum creatinine (sCr) per μ mol/L increase [33]. In another study that analyzed 167 confirmed patients with severe COVID-19, the LDH concentration was higher and the albumin concentration was lower in these patients, with significant differences [34]. Recently, Liang et al [22] used Least Absolute Shrinkage and Selection Operator (LASSO) and

logistic regression to construct a predictive risk score (COVID-GRAM), in which the AUC in the development cohort was 0.88 (95% CI 0.85-0.91) and the AUC in the validation cohort was 0.88 (95% CI 0.84-0.93) for predicting patients' risk of developing critical illness. The results in that study coincide with our results to some extent, and the machine learning algorithm can identify potential indicators better than the conventional algorithm. Finally, Abdulaal et al [20] used a neural network model to predict the prognosis of patients with COVID-19, with an AUC of 90.12%. This study only used eight indicators and achieved good predictive value; thus, it is more convenient and efficient.

Elevated LDH was identified as a significant risk predictive marker of COVID-19. Li et al [23] revealed that relatively high levels of LDH play a crucial role in predicting mortality of patients with COVID-19 when using an interpretable mortality prediction model. LDH is an enzyme that is involved in energy production through the conversion of lactate to pyruvate; it is present in almost all body cell types (n=156), with the highest levels in the heart, liver, lungs, muscles, kidneys, and blood cells. LDH is released from cells upon damage to their cytoplasmic membrane, and it is not only a metabolic marker but also an immune surveillance prognostic biomarker [35]. LDH increases the production of lactate, which leads to enhancement of immune-suppressive cells and inhibition of cytolytic cells. These changes weaken the immune response mounted against viral infection, which results in more severe disease in patients with COVID-19 who have elevated LDH [36]. PT is another typical indicator associated with patient prognosis. Through pathological examination of patients who died of COVID-19, researchers found that the virus can lead to disorders of the coagulation system, resulting in a hypercoagulable state and microthrombosis [37]. Moreover, viral infections may induce even more severe complications, such as acute respiratory distress syndrome and multi-organ dysfunction syndrome, which are two conditions frequently associated with hypercoagulation and disseminated intravascular coagulation [38]. These processes and conditions help to explain why PT was prolonged in patients with severe and critical COVID-19. Approximately 14.4% of patients with COVID-19 have elevated sCr levels, and kidney disease has been associated with in-hospital death of patients with COVID-19 [39]. SARS-CoV2 has been suggested to modulate the renin-angiotensin-aldosterone system (RAAS). Evidence of activation of the RAAS in patients with COVID-19 who have acute kidney injury, leading to increased sCr, has been reported [40]. Several studies have also reported that liver damage occurred in severe cases of COVID-19 infection at rates ranging from 58% to 78% [41,42]. COVID-19 uses angiotensin converting enzyme 2 (ACE2) as the binding site to enter host cells in the lungs, kidneys, and heart. A previous study [43] showed that both liver and bile duct cells express ACE2; this

may result in elevated T-Bil levels, accompanied by slightly decreased ALB levels. Hematologic and immunologic impairment showed significantly different profiles between survival and mortality of patients with COVID-19 with different disease severities. The results of our study suggest that increased NEUT% and decreased LYM% are risk factors for patient prognosis. Interestingly, a decrease in EOS% was also a risk factor, and we were surprised to find that the results of two studies [44,45] were consistent with our results. In those studies, it was found that eosinophils decreased at the early stage and were associated with disease severity and clinical outcomes. Impaired immune cell function leads to low lymphocyte levels and immune system dysfunction, causing patients with severe COVID-19 to be more sensitive to bacterial infection [46]. The decline in eosinophils may be due to the patients' response to the stress of acute SARS-CoV-2 infection. However, whether COVID-19 has a direct effect on eosinophils remains unknown. In one study, it was found that the *Clostridioides difficile* transferase toxin induces pathogenic host inflammation via a toll-like receptor 2–dependent pathway, resulting in suppression of the protective host eosinophilic response [47]. Additionally, eosinophils can be reduced after an innate immune challenge [48].

Limitations

This study has several limitations. First, data from only one center were used, and the sample size was small, which may indicate bias. However, Vulcan Hill Hospital is a large medical center that focused on the treatment of COVID-19. The patients are representative of all patients with COVID-19, providing a reliable basis for the treatment of critical patients. Second, the treatment of patients in the ICU is not necessarily the initial treatment because the patients were transferred from different hospitals and different medical treatment units; this may have affected the baseline characteristics of the patients. In the future, based on what we learned in this study, we will attempt to correct defects of the model and the machine learning approach. Also, we will collect more data to conduct external tests on the prediction model and further improve the generalized prediction ability of the model for multicenter data.

Conclusions

Machine learning has a good predictive effect on the mortality of critically ill patients with COVID-19 in the ICU. The XGBoost model has higher diagnostic performance than conventional statistical methods and can be used to select and simplify the core indicators for mortality prediction, such as LDH, PT, Cr, T-Bil, LYM%, ALB, and the white blood cell parameters NEUT% and BASO%. Machine learning may be a valuable prognostic indicator for early warning of critically ill patients; this warning plays a significant role in the allocation of medical resources, triage of patients, formulation of treatment decisions, and evaluation of progressive COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary table of the number and proportion of missing values.
[DOC File , 68 KB - jmir_v22i11e23128_app1.doc]

Multimedia Appendix 2
Mathematical algorithm.
[DOCX File , 13 KB - jmir_v22i11e23128_app2.docx]

Multimedia Appendix 3
Comparative analysis of all potential risk factors in the study cohort.
[DOC File , 200 KB - jmir_v22i11e23128_app3.doc]

Multimedia Appendix 4
Correlation coefficient matrix heat maps.
[DOCX File , 551 KB - jmir_v22i11e23128_app4.docx]

Multimedia Appendix 5
Load matrix diagram showing the correlations between the characteristics and the main factors.
[PNG File , 482 KB - jmir_v22i11e23128_app5.png]

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]

2. Zhen-Dong Y, Gao-Jun Z, Run-Ming J, Zhi-Sheng L, Zong-Qi D, Xiong X, et al. Clinical and transmission dynamics characteristics of 406 children with coronavirus disease 2019 in China: A review. J Infect 2020 Aug;81(2):e11-e15 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.030] [Medline: 32360500]

3. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]

4. Wander PL, Orlov M, Merel SE, Enquobahrie DA. Risk factors for severe COVID-19 illness in healthcare workers: Too many unknowns. Infect Control Hosp Epidemiol 2020 Apr 27:1-2 [FREE Full text] [doi: 10.1017/ice.2020.178] [Medline: 32336303]

5. Cen Y, Chen X, Shen Y, Zhang X, Lei Y, Xu C, et al. Risk factors for disease progression in patients with mild to moderate coronavirus disease 2019-a multi-centre observational study. Clin Microbiol Infect 2020 Jun 09:1242-1247 [FREE Full text] [doi: 10.1016/j.cmi.2020.05.041] [Medline: 32526275]

6. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]

7. Liu D, Wang Y, Wang J, Liu J, Yue Y, Liu W, et al. Characteristics and Outcomes of a Sample of Patients With COVID-19 Identified Through Social Media in Wuhan, China: Observational Study. J Med Internet Res 2020 Aug 13;22(8):e20108 [FREE Full text] [doi: 10.2196/20108] [Medline: 32716901]

8. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. J Med Internet Res 2020 Jun 30;22(6):e19636 [FREE Full text] [doi: 10.2196/19636] [Medline: 32544071]

9. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med 2020 Apr;26(4):506-510 [FREE Full text] [doi: 10.1038/s41591-020-0822-7] [Medline: 32284616]

10. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]

11. Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP Risk Grade: A Simple Mortality Index for Patients with Confirmed or Suspected Severe Acute Respiratory Syndrome Coronavirus 2 Disease (COVID-19) During the Early Stage of Outbreak in Wuhan, China. SSRN Journal Preprint posted online on February 28, 2020. [doi: 10.2139/ssrn.3543603]

12. Barda N, Riesel D, Akriv A, Levy J, Finkel U, Yona G, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun 2020 Sep 07;11(1):4439 [FREE Full text] [doi: 10.1038/s41467-020-18297-9] [Medline: 32895375]

13. Ewan CR, Bean D, Stammers M, Wang W, Zhang H, Searle T, et al. Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. medRxiv Preprint posted online on September 30, 2020. [doi: 10.1101/2020.04.24.20078006]

14. Cheng JW, Cao Y, Xu Z, Tan Z, Zhang X, Deng L, et al. Development and Evaluation of an AI System for COVID-19 Diagnosis. medRxiv Preprint posted online on June 02, 2020. [doi: 10.1101/2020.03.20.20039834]

15. Zhiyi WJ, Li Z, Hou R, Zhou L, Ye H, Chen Y, et al. Development and Validation of a Diagnostic Nomogram to Predict COVID-19 Pneumonia. medRxiv Preprint posted online on April 06, 2020. [doi: 10.1101/2020.04.03.20052068]

16. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 2020;20:100378 [FREE Full text] [doi: 10.1016/j.imu.2020.100378] [Medline: 32839734]

17. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and External Validation of a Prognostic Multivariable Model on Admission for Hospitalized Patients with COVID-19. SSRN Journal Preprint posted online on April 06, 2020. [doi: 10.2139/ssrn.3562456]

18. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, et al. Predicting COVID-19 Malignant Progression with AI Techniques. SSRN Journal Preprint posted online on March 31, 2020. [doi: 10.2139/ssrn.3557984]

19. Ying ZZ, Guo Y, Geng S, Gao S, Ye S, Hu Y, et al. A New Predictor of Disease Severity in Patients with COVID-19 in Wuhan, China. Research Square Preprint posted online on May 21, 2020. [doi: 10.21203/rs.3.rs-29566/v1]

20. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. J Med Internet Res 2020 Aug 25;22(8):e20259 [FREE Full text] [doi: 10.2196/20259] [Medline: 32735549]

21. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328. [doi: 10.1136/bmj.m1328] [Medline: 32265220]

22. Zampieri FG, Salluh JIF, Azevedo LCP, Kahn JM, Damiani LP, Borges LP, ORCHESTRA Study Investigators. ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis. Intensive Care Med 2019 Nov;45(11):1599-1607. [doi: 10.1007/s00134-019-05790-z] [Medline: 31595349]

23. Zack JE, Garrison T, Trovillion E, Clinkscale D, Coopersmith CM, Fraser VJ, et al. Effect of an education program aimed at reducing the occurrence of ventilator-associated pneumonia. Crit Care Med 2002 Nov;30(11):2407-2412. [doi: 10.1097/00003246-200211000-00001] [Medline: 12441746]

24. Lijiao ZJ, Liao M, Hua R, Huang P, Zhang M, Zhang Y, et al. Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study. medRxiv Preprint posted online on March 30, 2020. [doi: 10.1101/2020.03.25.20043166]

25. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 06;162(1):W1-73. [doi: 10.7326/M14-0698] [Medline: 25560730]

26. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med 2019 Jan 01;170(1):W1-W33. [doi: 10.7326/M18-1377] [Medline: 30596876]

27. de Oliveira Kaizer UA, Alexandre NMC, Rodrigues RCM, Cornélio ME, de Melo Lima MH, São-João TM. Measurement properties and factor analysis of the Diabetic Foot Ulcer Scale-short form (DFS-SF). Int Wound J 2020 Jun;17(3):670-682. [doi: 10.1111/iwj.13310] [Medline: 32065735]

28. Jung Y. Multiple predicting K-fold cross-validation for model selection. J Nonparametr Stat 2017 Nov 21;30(1):197-215. [doi: 10.1080/10485252.2017.1404598]

29. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput Aided Mol Des 2020 Oct 02;34(10):1013-1026 [FREE Full text] [doi: 10.1007/s10822-020-00314-0] [Medline: 32361862]

30. Meloni A, Maggio A, Positano V, Leto F, Angelini A, Putti MC, et al. CMR for myocardial iron overload quantification: calibration curve from the MIOT Network. Eur Radiol 2020 Jun;30(6):3217-3225. [doi: 10.1007/s00330-020-06668-1] [Medline: 32052169]

31. Palatnik de Sousa I, Maria Bernardes Rebuzzi Vellasco M, Costa da Silva E. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. Sensors (Basel) 2019 Jul 05;19(13):2969 [FREE Full text] [doi: 10.3390/s19132969] [Medline: 31284419]

XSL•FO
RenderX

32. Calculation tool for predicting the death probability of COVID-19 patients when entering the ICU. URL: http://114.251.235.51:1226/index [accessed 2020-10-27]

33. Chen L, Yu J, He W, Chen L, Yuan G, Dong F, et al. Risk factors for death in 1859 subjects with COVID-19. Leukemia 2020 Aug;34(8):2173-2183 [FREE Full text] [doi: 10.1038/s41375-020-0911-0] [Medline: 32546725]

34. Wei Y, Wang R, Zhang D, Tu Y, Chen C, Ji S, et al. Risk factors for severe COVID-19: Evidence from 167 hospitalized patients in Anhui, China. J Infect 2020 Jul;81(1):e89-e92 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.010] [Medline: 32305487]

35. Poggiali E, Zaino D, Immovilli P, Rovero L, Losi G, Dacrema A, et al. Lactate dehydrogenase and C-reactive protein as predictors of respiratory failure in CoVID-19 patients. Clin Chim Acta 2020 Oct;509:135-138 [FREE Full text] [doi: 10.1016/j.cca.2020.06.012] [Medline: 32531257]

36. Ding J, Karp JE, Emadi A. Elevated lactate dehydrogenase (LDH) can be a marker of immune suppression in cancer: Interplay between hematologic and solid neoplastic clones and their microenvironments. Cancer Biomark 2017 Jul 04;19(4):353-363. [doi: 10.3233/CBM-160336] [Medline: 28582845]

37. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. N Engl J Med 2020 Apr 23;382(17):e38 [FREE Full text] [doi: 10.1056/NEJMc2007575] [Medline: 32268022]

38. Chang JC. Acute Respiratory Distress Syndrome as an Organ Phenotype of Vascular Microthrombotic Disease: Based on Hemostatic Theory and Endothelial Molecular Pathogenesis. Clin Appl Thromb Hemost 2019;25:1076029619887437. [doi: 10.1177/1076029619887437] [Medline: 31775524]

39. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020 May;97(5):829-838 [FREE Full text] [doi: 10.1016/j.kint.2020.03.005] [Medline: 32247631]

40. Dudoignon E, Moreno N, Deniau B, Coutrot M, Longer R, Amiot Q, et al. Activation of the renin-angiotensin-aldosterone system is associated with Acute Kidney Injury in COVID-19. Anaesth Crit Care Pain Med 2020 Aug;39(4):453-455 [FREE Full text] [doi: 10.1016/j.accpm.2020.06.006] [Medline: 32565254]

41. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]

42. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. Am J Gastroenterol 2020 May;115(5):766-773 [FREE Full text] [doi: 10.14309/ajg.0000000000000620] [Medline: 32287140]

43. Parohan M, Yaghoubi S, Seraji A. Liver injury is associated with severe coronavirus disease 2019 (COVID-19) infection: A systematic review and meta-analysis of retrospective studies. Hepatol Res 2020 Aug;50(8):924-935 [FREE Full text] [doi: 10.1111/hepr.13510] [Medline: 32386449]

44. Lu G, Wang J. Dynamic changes in routine blood parameters of a severe COVID-19 case. Clin Chim Acta 2020 Sep;508:98-102 [FREE Full text] [doi: 10.1016/j.cca.2020.04.034] [Medline: 32405079]

45. Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, Medical Treatment Expert Group for COVID-19. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. J Allergy Clin Immunol 2020 Jul;146(1):89-100 [FREE Full text] [doi: 10.1016/j.jaci.2020.05.003] [Medline: 32407836]

46. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]

47. Cowardin CA, Buonomo EL, Saleh MM, Wilson MG, Burgess SL, Kuehne SA, et al. The binary toxin CDT enhances Clostridium difficile virulence by suppressing protective colonic eosinophilia. Nat Microbiol 2016 Jul 11;1(8):16108 [FREE Full text] [doi: 10.1038/nmicrobiol.2016.108] [Medline: 27573114]

48. Hassani M, Leijte G, Bruse N, Kox M, Pickkers P, Vrisekoop N, et al. Differentiation and activation of eosinophils in the human bone marrow during experimental human endotoxemia. J Leukoc Biol 2020 Nov;108(5):1665-1671. [doi: 10.1002/JLB.1AB1219-493R] [Medline: 31922294]

Abbreviations

ACE2: angiotensin converting enzyme 2
AdaBoost: adaptive boosting
ALB: albumin level
APACHE II: Acute Physiologic Assessment and Chronic Health Evaluation II
AUC: area under the receiver operating characteristic curve
BASO%: basophil percentage
Cr: creatinine
EOS%: eosinophil percentage
GBDT: gradient boosting decision tree

ICU: intensive care unit
LASSO: Least Absolute Shrinkage and Selection Operator
LDH: lactate dehydrogenase
LYM%: lymphocyte percentage
MODS: Multiple Organ Disfunction Score
NEUT%: neutrophil percentage
NLR: negative likelihood ratio
NPV: negative predictive value
PLR: positive likelihood ratio
PPV: positive predictive value
PROBAST: Prediction model Risk Of Bias ASsessment Tool
PSI: Pneumonia Severity Index
PT: prothrombin time
RAAS: renin-angiotensin-aldosterone system
sCr: serum creatinine
SHAP: SHapley Additive exPlanations
SOFA: Sequential Organ Failure Assessment
T-Bil: total bilirubin
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
XGBoost: eXtreme Gradient Boosting

XSL•FO
RenderX

Impact of Systematic Factors on the Outbreak Outcomes of the Novel COVID-19 Disease in China: Factor Analysis Study

Zicheng Cao[1*]; Feng Tang[1*]; Cai Chen[1]; Chi Zhang[1]; Yichen Guo[1]; Ruizhen Lin[1]; Zhihong Huang[1]; Yi Teng[1]; Ting Xie[1]; Yutian Xu[2]; Yanxin Song[3]; Feng Wu[1]; Peipei Dong[1]; Ganfeng Luo[1]; Yawen Jiang[1], PhD; Huachun Zou[1], PhD; Yao-Qing Chen[1], PhD; Litao Sun[1], PhD; Yuelong Shu[1], PhD; Xiangjun Du[1], PhD

[1]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou, China

[2]School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China

[3]Lingnan College, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Xiangjun Du, PhD
School of Public Health (Shenzhen)
Sun Yat-sen University
135 Xingang Xi Road
Guangzhou, 510275
China
Phone: 86 020 83226383
Email: duxj9@mail.sysu.edu.cn

Abstract

Background: The novel COVID-19 disease has spread worldwide, resulting in a new pandemic. The Chinese government implemented strong intervention measures in the early stage of the epidemic, including strict travel bans and social distancing policies. Prioritizing the analysis of different contributing factors to outbreak outcomes is important for the precise prevention and control of infectious diseases. We proposed a novel framework for resolving this issue and applied it to data from China.

Objective: This study aimed to systematically identify national-level and city-level contributing factors to the control of COVID-19 in China.

Methods: Daily COVID-19 case data and related multidimensional data, including travel-related, medical, socioeconomic, environmental, and influenza-like illness factors, from 343 cities in China were collected. A correlation analysis and interpretable machine learning algorithm were used to evaluate the quantitative contribution of factors to new cases and COVID-19 growth rates during the epidemic period (ie, January 17 to February 29, 2020).

Results: Many factors correlated with the spread of COVID-19 in China. Travel-related population movement was the main contributing factor for new cases and COVID-19 growth rates in China, and its contributions were as high as 77% and 41%, respectively. There was a clear lag effect for travel-related factors (previous vs current week: new cases, 45% vs 32%; COVID-19 growth rates, 21% vs 20%). Travel from non-Wuhan regions was the single factor with the most significant impact on COVID-19 growth rates (contribution: new cases, 12%; COVID-19 growth rate, 26%), and its contribution could not be ignored. City flow, a measure of outbreak control strength, contributed 16% and 7% to new cases and COVID-19 growth rates, respectively. Socioeconomic factors also played important roles in COVID-19 growth rates in China (contribution, 28%). Other factors, including medical, environmental, and influenza-like illness factors, also contributed to new cases and COVID-19 growth rates in China. Based on our analysis of individual cities, compared to Beijing, population flow from Wuhan and internal flow within Wenzhou were driving factors for increasing the number of new cases in Wenzhou. For Chongqing, the main contributing factor for new cases was population flow from Hubei, beyond Wuhan. The high COVID-19 growth rates in Wenzhou were driven by population-related factors.

Conclusions: Many factors contributed to the COVID-19 outbreak outcomes in China. The differential effects of various factors, including specific city-level factors, emphasize the importance of precise, targeted strategies for controlling the COVID-19 outbreak and future infectious disease outbreaks.

KEYWORDS

COVID-19; new cases; growth rate; multidimensional factors; statistical machine learning

Introduction

A new pneumonia disease emerged and was later named COVID-19 [1]. COVID-19 is caused by the novel SARS-CoV-2 and has become a major global health threat [2,3]. Due to massive population movement during the early stage of the COVID-19 epidemic, the disease rapidly spread in China. To keep the spread of COVID-19 in control, the Chinese government implemented rapid and strict intervention measures, such as quarantining Wuhan and its surrounding cities in the Hubei Province on January 23 and 24, 2020, banning public transportation, cancelling activities involving a gathering of people, extending the Spring Festival holiday, postponing the opening of schools, and setting up monitoring, testing, and isolation policies [4,5]. It has been proven that these strong measures have effectively slowed down, and even prevented, the spread of COVID-19 in China. The total number of new cases across the country has decreased rapidly and has been kept at a low level in the past several months [6,7].

The spread of COVID-19 depends on many factors. The human population is naïve to SARS-CoV-2, but there is currently no evidence showing dramatic changes in the virus [8]. Until now, many studies based on both statistical and mechanistic models have explored and confirmed the effect of population movement on the spread of COVID-19 [9-11]. Socioeconomic status, climate conditions, and intervention measures vary among cities across China. As a result, regional systems, resources, and the country's capacity for responding to public health risks and events are directly linked to the outcomes of an outbreak [12]. Moreover, the allocation of medical resources and other related factors, such as the number of hospital beds, have a positive effect on the control of the epidemic [13,14]. In addition, many socioeconomic factors, such as population number, population density, and social activities, mediate the spread of the disease [15]. Studies have confirmed that environmental factors influence the seasonal transmission of pathogens, but the effect of environmental factors on COVID-19 is still controversial [16-19]. Furthermore, while previous upper tract respiratory infections might relate to infections by human coronaviruses and provide some cross-protection against SARS-CoV-2, the relationship between previous upper tract respiratory infections and COVID-19 is unclear and must be further studied [20]. Therefore, although the integrated effects of several influencing factors on the outbreak outcomes of COVID-19 have been analyzed [21], a comprehensive analysis has yet to be conducted.

We collected a comprehensive dataset and used a correlation analysis and machine learning algorithm to identify and assess national-level and city-level contributing factors to the outbreak outcomes of COVID-19 in China.

Methods

Data

Data Collection

As of March 1, 2020, the number of new COVID-19 cases in most prefecture-level cities in China has declined to 0. Therefore, data from 343 prefecture-level cities in China from January 17 to February 29, 2020 were collected and used in this study. This period was also the main COVID-19 outbreak period in China. Data were evaluated on a weekly scale comprised of the following 6 weeks: week 1, January 17-23; week 2, January 24-30; week 3, January 31 to February 6; week 4, February 7-13; week 5, February 14-20; and week 6, February 21-29. For each week, cities with 0 new cases were excluded from further analysis. Weekly new cases and COVID-19 growth rates were considered response/dependent variables, and a variety of regressors/independent variables were included in this study. These independent variables were divided into the following 6 categories: travel-related (current week), travel-related (previous week), medical, socioeconomic, environmental, and influenza-like illness (ILI) variables (Table S1 in Multimedia Appendix 1).

New Cases and COVID-19 Growth Rates

Confirmed COVID-19 case data were downloaded from daily official reports from the health commission, and weekly accumulative new case data were extracted. A proxy measurement for the reproductive number (R_{proxy}) was used as the indicator for COVID-19 growth rate, which was defined as the number of new cases in the following week normalized by the number of new cases in the current week [22], as follows:

In this equation, w denotes week, ranging from 1-6, N is the number of weekly cumulative new cases, and i represents the ith city.

Travel-Related Factors

The daily domestic population movement data were derived from Baidu Qianxi [23]. The data are based on the positioning and transportation information systems of Baidu Location Based Services and Baidu Tianyan. The system collects location information that is voluntarily uploaded by users using Baidu services in real time and draws a population migration map based on global positioning system, Wi-Fi location, IP address, and signal tower information. We obtained city-level crowd movement information from this map for travel-related data. The following 5 city-level measurements were used in this study: population flow from Wuhan (Wuhan flow), population flow from Hubei, excluding Wuhan (Hubei/non-Wuhan flow), population flow from regions in mainland China, excluding Hubei (non-Hubei flow), population flow within a city (city internal flow), and the activity intensity within a city (city internal flow index), which was obtained by normalizing the

population flow within a city by its population. Weekly measures were obtained based on the sum of daily data. Due to the latent period of COVID-19, travel-related factors from the previous week were considered separate from travel-related factors from the current week in this study. Hubei/non-Wuhan flow and non-Hubei flow measurements were combined to obtain non-Wuhan flow measurements, and city internal flow and the city internal flow index were combined to obtain city flow measurements.

Medical Factors

The number of doctors, hospitals, beds, and outpatients and emergency patients were derived from the 2018 edition of the China Health & Family Planning Statistical Yearbook [24]. Only province-level data were available for outpatient data, so city-level values were obtained in proportion to each city's population. Data for the number of COVID-19 treatment hospitals were extracted from announcements issued by provincial health committees. Values were kept the same across all 6 weeks for each city.

Socioeconomic Factors

Socioeconomic factors, including population number, population density, gross domestic product, per capita income, and percentage of the population aged ≥65 years were derived from the 2018 China City Statistical Yearbook [25]. Except for the percent of the population aged ≥65 years, all socioeconomic data were at the city level. Province-level data for the percent of the population aged ≥65 years were used for each city. Values were kept the same across all 6 weeks for each city.

Environmental Factors

The daily climate data for each city, which included the highest temperature, lowest temperature, average temperature, relative humidity, and absolute humidity were downloaded from the China Meteorological Administration website [26]. Weekly data were calculated by averaging the daily data.

ILI factors

The average percentage of ILI occurrences for each city from 2016 to 2018 were calculated based on weekly report data from the Chinese National Influenza Center. The data were based on the number of samples tested in 554 sentinel hospitals in 31 provinces in China.

Correlation Analysis

Spearman correlation was used to evaluate the relationship between a single factor and either the number of new cases or COVID-19 growth rates. A significance level of .05 was used in this study.

Machine Learning Framework

Feature Selection and Feature Importance

First, a nonlinear regression tree model, which was made with XGBoost framework (extreme gradient boosting) [27], was used to fit weekly new cases and COVID-19 growth rates. Next, sequential backward floating selection was iterated to train the XGBoost [28,29] model to obtain the final model by minimizing the mean squared error. Sequential backward floating selection

is a sequential feature selection method based on a greedy search algorithm. It removes features one by one from the full set of features and evaluates the error function. When the error reaches the optimum level, the combination of left-over features is regarded as the optimal feature combination. Finally, the importance of each selected factor was determined based on the number of times the factor was split by the tree model, which was determined by XGBoost. For each week, a contribution percentage was calculated for each factor based on its relative importance [30].

Shapely Additive Explanation Analysis

Shapley additive explanation (SHAP) [31] is an interpretable method for analyzing the output of machine learning models. SHAP analysis was used to calculate the contribution of selected factors for each city and week. A predicted value (ie, SHAP value) was generated for each data sample, and the value was uniformly assigned to each feature of the sample. The following equation shows how the predicted value of the machine learning model was calculated:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \ldots + f(x_{iN}) \qquad (2)$$

In this equation, the ith sample was defined as x_i, the jth feature of the ith sample was defined as x_{ij}, the predicted value of the machine learning model for the sample was y_i, and the reference value of the model (ie, the mean value of the target sample variable) was defined as y_{base}. Furthermore, $f(x_{ij})$ was the SHAP value of x_{ij}. $f(x_{ij})>0$ indicates that the feature increases the predicted value and has a positive contribution; otherwise, the feature reduces the predicted value and has a negative contribution.

SHAP package [32] was used to calculate a marginal contribution value (ie, SHAP value) based on the trained XGBoost model [33-35]. The SHAP value was used to measure the contribution of different important factors for each city.

Results

Single-Factor Analysis

Based on the single-factor Spearman correlation analysis, many factors significantly correlated with both the number of new cases and COVID-19 growth rates in China (Figure 1). All factors significantly correlated with the number of new cases during at least 1 week, except relative humidity. Similarly, only travel-related (current and previous weeks), medical, and socioeconomic factors significantly correlated with COVID-19 growth rates during at least one week, while environmental and ILI factors did not correlate with COVID-19 growth rates (Figure 1). Wuhan flow, Hubei/non-Wuhan flow, and their corresponding measures in previous weeks positively correlated with the number of new cases during all 6 weeks, while the city internal flow index negatively correlated with the number of new cases during all 6 weeks. Many factors significantly correlated with the number of new cases for at least 5 weeks, but only Wuhan flow from the previous week and population density significantly correlated with COVID-19 growth rates for at least 5 weeks (Figure 1).

Figure 1. Correlation analysis of contributing factors for new COVID-19 cases (left) and COVID-19 growth rates (right). Dots from the top to the bottom for each factor indicate the corresponding Spearman correlation coefficient for 6 consecutive weeks (January 17 to February 29, 2020). Significant correlations are colored in red. ILI: influenza-like illness.
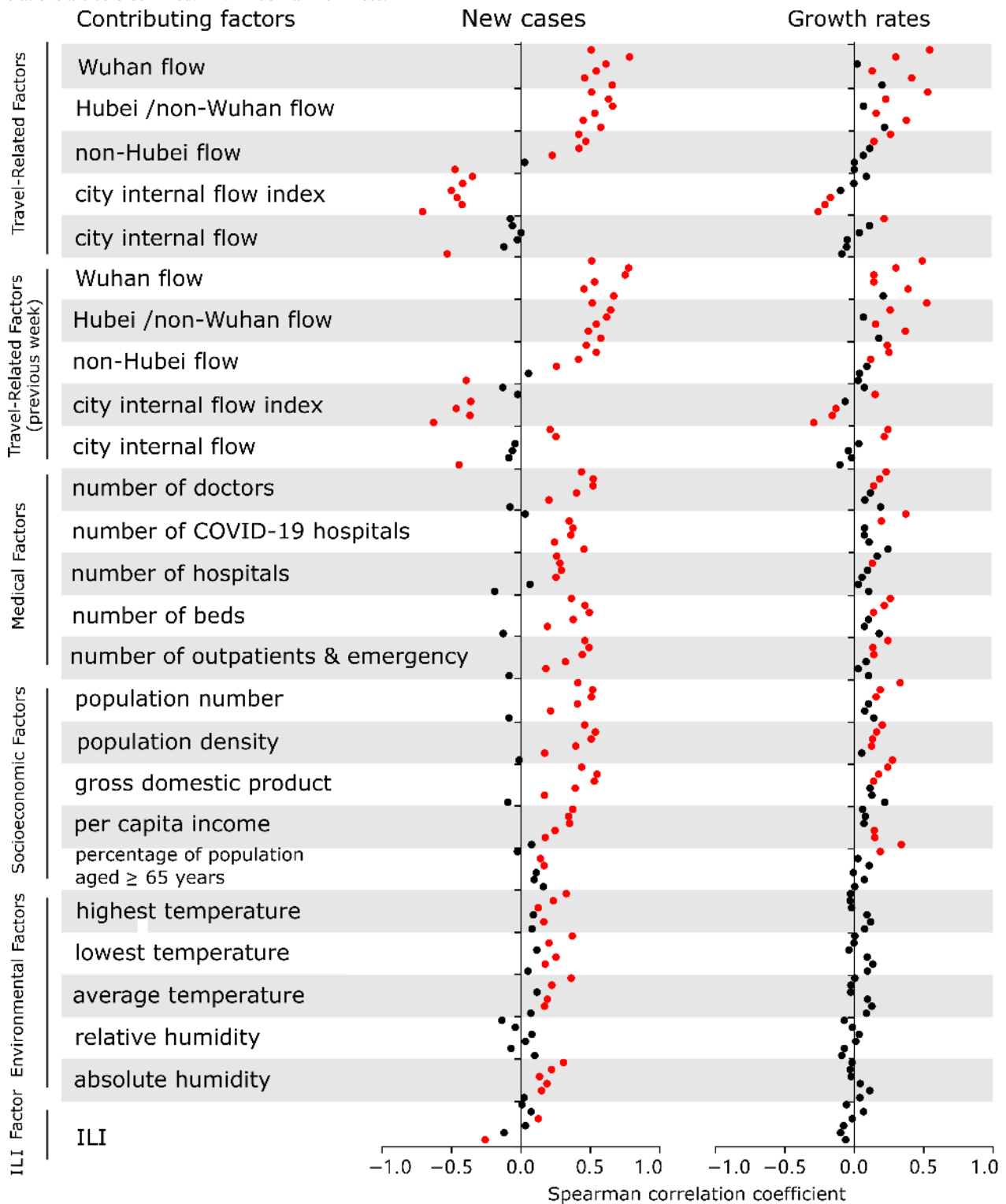
Multifactor Analysis Based on the Machine Learning Algorithm

Our additional analysis showed that many factors correlated with each other (Figure S1 in Multimedia Appendix 2). As a result, a nonlinear method was needed to assess the contribution of correlated factors on the spread of COVID-19 in China. Machine learning methods are good for solving nonlinear

problems, and we used XGBoost [28,30] to create a nonlinear regression tree model for this study. Important factors were selected based on the cross-validation procedure in the XGBoost framework. Figure S2 in Multimedia Appendix 2 shows the number of factors selected for the final model. No single factor was consistently selected for all 6 weeks, but the overall contribution percentage could be obtained (Figure 2, Multimedia Appendix 2). Figure 2 shows that travel-related factors for the

current week (contribution: new cases, 32%; COVID-19 growth rates, 20%) and related measures from the previous week (contribution: new cases, 45%; COVID-19 growth rates, 21%) were the main factors that drove changes in the number of new cases and COVID-19 growth rates, with total contribution percentages of 77% and 41%, respectively. For new cases, the contributions of other factors were no more than 10%. For COVID-19 growth rates, socioeconomic factors were also important, and they had a combined contribution percentage of 28%. Other factors also contributed to COVID-19 growth rates in China, with environmental, ILI, and medical factors contributing 14%, 10%, and 7%, respectively. For new cases, the leading individual factors with contribution percentages >10% were Wuhan flow, city flow, and non-Wuhan flow, with combined contributions (ie, travel-related measures and corresponding measures from previous week) of 49%, 16%, and 12%, respectively. For COVID-19 growth rates, the leading contributing factors were non-Wuhan flow, population density, ILIs, and Wuhan flow, with combined contributions of 26%, 10%, 10%, and 9%, respectively. The remaining factors with contributions of no less than 5% for the growth rate of COVID-19 were city flow (current week and previous week), population number, the number of COVID-19 hospitals, lowest temperature, absolute humidity, and per capita income, with contributions of 7%, 6%, 6%, 6%, 6%, and 5%, respectively (Figure 2).

Figure 2. Contribution of various factors to new COVID-19 cases and COVID-19 growth rates in China. ILI: influenza-like illness.
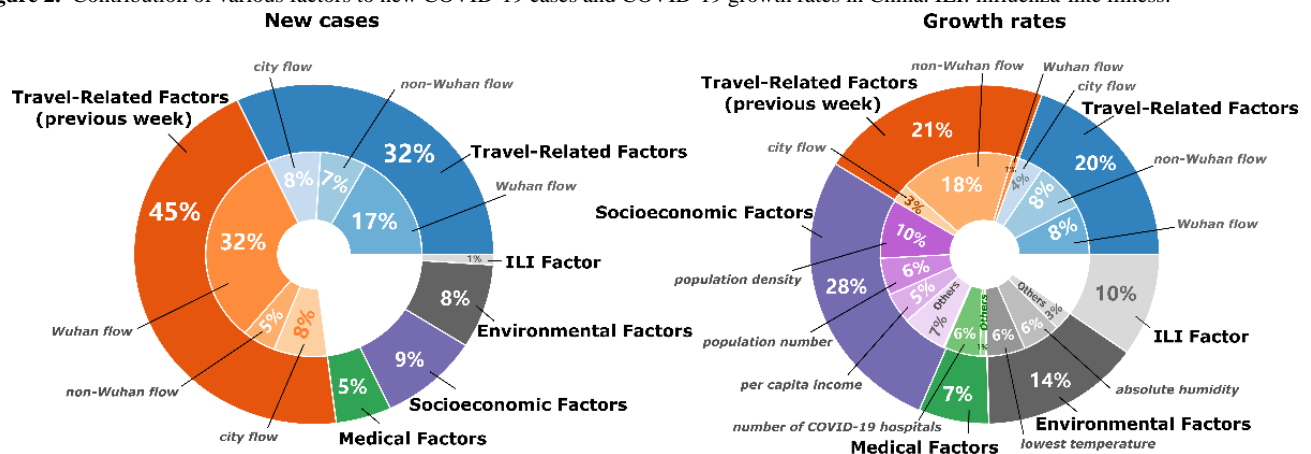
City-Level Factor Analysis

SHAP is a method for explaining individual predictions [31,35] by computing the contribution of each factor and comparing it to predictions based on game theory. Multimedia Appendix 3 shows the detailed SHAP values of each selected factor for each city. SHAP values explain the contribution of different factors to the COVID-19 outbreak outcomes of each individual city. For a demonstration, we selected a number of new cases from the second week (106, 221, and 179, respectively) and COVID-19 growth rates from the first week (4.08, 36.83, and 6.63, respectively) from Beijing, Wenzhou, and Chongqing (Figure 3). In the second week, Wuhan flow and city internal flow (previous week) contributed to an increased number of new cases in Wenzhou (overall SHAP value, 90.75), while Hubei/non-Wuhan flow was the main factor that led to an increased number of new cases in Chongqing (SHAP value, 118.66) (Figure 3A). For Beijing, population density and Wuhan flow were the top 2 factors that contributed to the number of new cases, with a combined SHAP value of 41.29, which was much smaller than the SHAP values for factors in Wenzhou and Chongqing. This also meant that factors in Beijing related to fewer new cases than factors in Wenzhou and Chongqing (ie, 106 new cases vs 221 and 179 new cases, respectively) (Figure 3A). Furthermore, the differences in COVID-19 growth rates in Wenzhou compared to those in Beijing and Chongqing in the first week (ie, 36.83 vs 4.08 and 6.63, respectively) were caused by population number, non-Hubei flow, per capita income, and the percentage of the population aged ≥65 years (SHAP value: 10.87, 1.67, 1.64, and 1.48, respectively) (Figure 3B). Among them, the contribution from population number was positive for all 3 cities and bigger for Wenzhou, while non-Hubei flow and per capita income were positive for Wenzhou and negative for the other cities, which indicates the importance of those factors to the higher COVID-19 growth rates in Wenzhou. For Beijing, the contribution of the number of COVID-19 hospitals was bigger than in Wenzhou and Chongqing (SHAP value: 6.06 vs 0.93 and 1, respectively) (Figure 3B).

Figure 3. SHAP values for selected factors in Beijing, Wenzhou, and Chongqing. (A) New cases of COVID-19 in the second week. (B) Growth rates of COVID-19 in the first week. Factors were selected by the nonlinear regression tree model. SHAP: Shapely additive explanation.
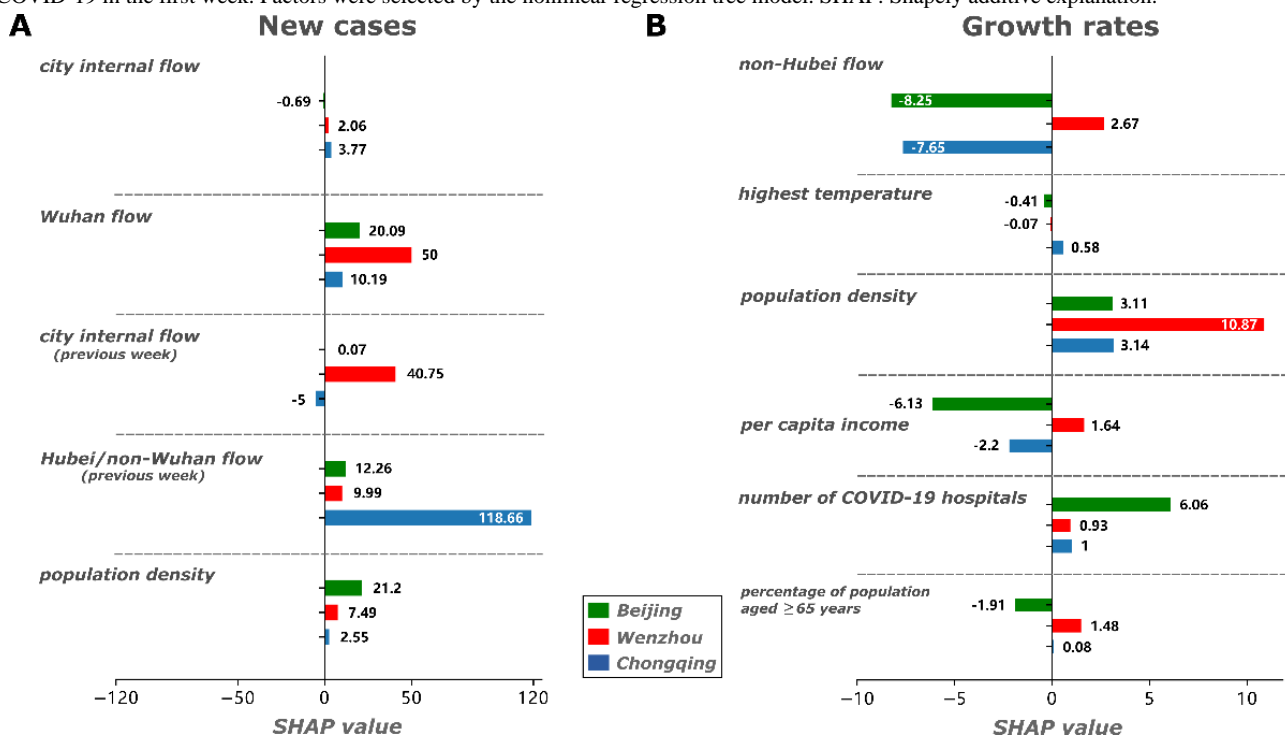
Discussion

Principal Results

In this study, we applied a machine learning method to delineate the contribution of different factors to weekly new cases and COVID-19 growth rates based on multidimensional data collected from 343 cities in China. Travel-related factors were the main contributors to the number of new cases and COVID-19 growth rates in China during the outbreak period (ie, January 17 to February 29, 2020), and the growth rate was also affected by other factors, including socioeconomic factors like population density [36]. City-level differences among related factors led to different COVID-19 outbreak outcomes and shed light on targeted prevention and control methods for the future.

The results of our simple correlation analysis and machine learning exploration showed that the progression of the COVID-19 outbreak in China is affected by multiple factors. Based on our nonlinear machine learning method, we found that the most important contributors to new cases and COVID-19 growth rates in China were travel-related factors and that travel-related factors had a clear lag effect that could not be ignored. Previous studies have shown that population movement from Wuhan is the driving factor for new COVID-19 cases in China, and we have confirmed this [2,37,38]. Additionally, we found that population movement from regions beyond Wuhan and internal flow within the city significantly contributed to new cases and COVID-19 growth rates in China, and the population flow from non-Wuhan regions was the leading contributor for COVID-19 growth rates (Figure 2). City internal flow also contributed to the increase in COVID-19 cases and spread of the disease, but its impact was limited. The significant contribution of population movement from non-Wuhan regions

emphasizes the importance of monitoring and restricting population movement from regions beyond the disease epicenter during an outbreak. This may relate to the fact that, for people to successfully return to their hometowns and avoid trouble, they might travel in a circuitous manner. Therefore, authorities should pay attention to all travelers, not only the ones directly from Wuhan, especially considering the lag effect of population movement.

Our study also indicated that the contribution of other factors to the progression of the COVID-19 outbreak in China cannot be ignored, especially for the growth rate of COVID-19. The contribution of socioeconomic factors to the growth rate of COVID-19 in China is comparable to that of travel-related factors [39] (Figure 2). The leading contributing socioeconomic factor was population density; a higher population density means a higher probability of secondary infection, resulting in the faster growth of new COVID-19 cases. Among all the factors considered in this study, factors related to medical resources contributed the least to new COVID-19 cases in China, which may indicate that there were enough medical resources for most cities in China. Interestingly, the dominant contributing factor for COVID-19 growth rates among all medical factors was the number of COVID-19 hospitals. This factor had a contribution of 6%, which corresponds to more than 85% of the contribution of all medical factors, indicating that the practice of setting up designated hospitals for COVID-19 in various cities is effective for controlling the growth of cases. Our results also indicated that environmental and ILI factors contributed to COVID-19 growth rates in China, but their contributions were smaller compared to those of travel-related and socioeconomic factors. Therefore, the contribution of environmental and ILI factors should be interpreted carefully and studied further [40,41].

Based on the SHAP values used in game theory, we were able to distinguish the individual contribution of different city-level factors, which has important implications for precise and targeted control strategies. For example, compared to the number of new cases in Beijing, which, as the capital city, is a super megacity with a large population and a hub of population movement, Wenzhou and Chongqing had more new cases in the second week with different contributing factors. Wuhan flow and city internal flow (previous week) were the main contributors for increases in new COVID-19 cases in Wenzhou, while Hubei/non-Wuhan flow (previous week) was the driving factor in Chongqing. To reduce the number of new COVID-19 cases in Wenzhou, efforts beyond restricting population migration from Wuhan are needed, such as reducing the number of social activities within the city. As the adjacent city of Hubei, Chongqing should pay more attention to travelers from Hubei. With regard to the higher COVID-19 growth rates during the first week in Wenzhou, beyond travel-related factors, contributing factors were mainly socioeconomic factors, including population number, per capita income, and percentage of the population aged ≥65 years. Ours is the first study to evaluate the contribution of different city-level factors to outbreak outcomes, and our results and methodology are helpful for the targeted control of infectious diseases.

Limitations

This study had several limitations. First, although we assessed as many factors as possible, important factors might be missing. For example, although we included measures related to the social distancing policy, such as the city internal flow and city internal flow index, other detailed control policies [42] are missing. Second, due to the multicollinearity among factors and black box effect in the model, our results may not be biologically sound and require careful interpretation. Therefore, negative SHAP values should not be interpreted as factors with a negative effect. Instead, negative SHAP values can indicate a compromised effect and small contribution. Third, more data on medical resources are needed in the future, especially data on the redistribution of medical resources during the epidemic, which are important for effective future resource arrangement. Fourth, although we used a powerful machine learning method to deal with the complex relationships between different factors, it may not the most suitable method. Therefore, other methods should be explored, especially since the data in this study cover a short period of time and exhibit dramatic changes due to strong interventions. Nevertheless, our quantitative results and proposed method shed light on the contribution of different factors to outbreak outcomes and are useful for the precise prevention and control of infectious diseases.

Conclusions

The prevention and control of the COVID-19 epidemic is a systematical project. Knowing the important contributing factors and prioritizing the corresponding strategies are helpful for creating effective control measures. Beyond population flow from Wuhan, population flow from other places and internal flow within the city also contributed to the number of new cases and COVID-19 growth rates in China. Socioeconomic factors, particularly population number and density, also play very important roles in COVID-19 growth rates in China. The contribution of specific factors for individual cities was also explored based on the framework proposed in this study. The pandemic is still ongoing worldwide, and many countries are experiencing the severe rebound effects of COVID-19. The results we presented and the framework we proposed in this study are helpful and useful for exploring optimal and precise control strategies.

Authors' Contributions

XD and ZC designed the study. ZC, FT, CC, CZ, YG, RL, ZH, YT, TX, YXu, YS, FW, PD, and GL collected and analyzed the data. XD, YJ, HZ, YC LS, and YShu interpreted the data. XD, ZC, and FT prepared the manuscript. XD, GL, YJ, HZ, YC, LS, and YShu edited the paper. All authors reviewed and approved the submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. A collection of all influencing factors.
[ZIP File (Zip Archive), 131 KB - jmir_v22i11e23853_app1.zip]

Multimedia Appendix 2
Supplementary materials, including Figure S1, Figure S2 and Table S2.

XSL•FO

RenderX

[DOCX File , 704 KB - jmir_v22i11e23853_app2.docx]

Multimedia Appendix 3
Shapley additive explanation values of each selected factor for each city.
[ZIP File (Zip Archive), 32 KB - jmir_v22i11e23853_app3.zip]

References

1. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]

3. World Health Organization. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 2020-06-09]

4. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott D. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]

5. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]

6. Li Z, Chen Q, Feng L, Rodewald L, Xia Y, Yu H. Active case finding with case management: the key to tackling the COVID-19 pandemic. Lancet 2020 Jul 04;396(10243):63-70 [FREE Full text] [doi: 10.1016/S0140-6736(20)31278-2] [Medline: 32505220]

7. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020 Mar 07;395(10226):764-766 [FREE Full text] [doi: 10.1016/S0140-6736(20)30421-9] [Medline: 32105609]

8. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]

9. Anzai A, Kobayashi T, Linton NM, Kinoshita R, Hayashi K, Suzuki A, et al. Assessing the Impact of Reduced Travel on Exportation Dynamics of Novel Coronavirus Infection (COVID-19). J Clin Med 2020 Feb 24;9(2):601 [FREE Full text] [doi: 10.3390/jcm9020601] [Medline: 32102279]

10. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]

11. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for Transportation of Coronavirus Disease from Wuhan to Other Cities in China. Emerg Infect Dis 2020 May;26(5):1049-1052 [FREE Full text] [doi: 10.3201/eid2605.200146] [Medline: 32053479]

12. Tam T. Fifteen years post-SARS: Key milestones in Canada's public health emergency response. Can Commun Dis Rep 2018 May 03;44(5):98-101 [FREE Full text] [doi: 10.14745/ccdr.v44i05a01] [Medline: 31007618]

13. Kandel N, Chungong S, Omaar A, Xing J. Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries. Lancet 2020 Mar 28;395(10229):1047-1053 [FREE Full text] [doi: 10.1016/S0140-6736(20)30553-5] [Medline: 32199075]

14. Liang L, Tseng C, Ho HJ, Wu C. Covid-19 mortality is negatively associated with test number and government effectiveness. Sci Rep 2020 Jul 24;10(1):12567 [FREE Full text] [doi: 10.1038/s41598-020-68862-x] [Medline: 32709854]

15. Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J Popul Econ 2020 May 09:1-46 [FREE Full text] [doi: 10.1007/s00148-020-00778-2] [Medline: 32395017]

16. Price RHM, Graham C, Ramalingam S. Association between viral seasonality and meteorological factors. Sci Rep 2019 Jan 30;9(1):929 [FREE Full text] [doi: 10.1038/s41598-018-37481-y] [Medline: 30700747]

17. Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ 2020 Jul 01;724:138201 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138201] [Medline: 32408450]

18. Qu G, Li X, Hu L, Jiang G. An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19). Environ Sci Technol 2020 Apr 07;54(7):3730-3732 [FREE Full text] [doi: 10.1021/acs.est.0c01102] [Medline: 32202420]

19. Yao Y, Pan J, Liu Z, Meng X, Wang W, Kan H, et al. No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. Eur Respir J 2020 May;55(5):2000517. [doi: 10.1183/13993003.00517-2020] [Medline: 32269084]

20. Kong W, Li Y, Peng M, Kong D, Yang X, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. Nat Microbiol 2020 May;5(5):675-678. [doi: 10.1038/s41564-020-0713-1] [Medline: 32265517]

XSL•FO
RenderX

21. Ran J, Zhao S, Han L, Liao G, Wang K, Wang MH, et al. A re-analysis in exploring the association between temperature and COVID-19 transmissibility: an ecological study with 154 Chinese cities. Eur Respir J 2020 Aug;56(2):2001253. [doi: 10.1183/13993003.01253-2020] [Medline: 32631839]

22. Luo W, Majumder M, Liu D, Poirier C, Mandl KD, Lipsitch M, et al. The role of absolute humidity on transmission rates of the COVID-19 outbreak. medRxiv Preprint posted online on February 17, 2020. [doi: 10.1101/2020.02.12.20022467]

23. Baidu Qianxi. Baidu Migration. URL: http://qianxi.baidu.com/ [accessed 2020-11-09]

24. China Health and Family Planning Statistical Yearbook 2018. China Data Insights. URL: http://cdi.cnki.net/Titles/SingleNJ?NJCode=N2019030282 [accessed 2020-06-09]

25. National Bureau of Statistics. 2018 China City Statistical Yearbook. URL: http://www.stats.gov.cn/tjsj/tjcbw/201907/t20190708_1674721.html [accessed 2020-06-09]

26. China Meteorological Administration. Weather Forecast. URL: https://tianqi.911cha.com/ [accessed 2020-11-09]

27. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Aug Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2016; San Francisco, California p. 785-794. [doi: 10.1145/2939672.2939785]

28. Ferri FJ, Pudil P, Hatef M, Kittler C. Comparative study of techniques for large-scale feature selection. Machine Intelligence and Pattern Recognition 1994;16:403-413 [FREE Full text] [doi: 10.1016/B978-0-444-81892-8.50040-7]

29. Raschka S. MLxtend: Providing machine learning and data scienceutilities and extensions to Python's scientific computingstack. J Open Source Softw 2018 Apr;3(24):638. [doi: 10.21105/joss.00638]

30. Shi X, Wong YD, Li MZ, Palanisamy C, Chai C. A feature learning approach based on XGBoost for driving assessment and risk prediction. Accid Anal Prev 2019 Aug;129:170-179. [doi: 10.1016/j.aap.2019.05.005] [Medline: 31154284]

31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017; Long Beach, CA.

32. GitHub. SHAP Package. URL: https://github.com/slundberg/shap [accessed 2020-11-09]

33. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using improved Shapley Additive Explanation. In: Explanation of Machine Learning Models Using Improved Shapley Additive Explanation. 2019 Sep Presented at: BCB '19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; September 2019; Niagara Falls, NY p. 546. [doi: 10.1145/3307339.3343255]

34. Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing. PeerJ Comput Sci 2017 Jul 24;3:e127. [doi: 10.7717/peerj-cs.127]

35. Lundberg SM, Erion GG, Lee SI. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv Preprint posted online on March 7, 2019.

36. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 2020 Jun;582(7812):389-394. [doi: 10.1038/s41586-020-2284-y] [Medline: 32349120]

37. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]

38. Zhao S, Zhuang Z, Cao P, Ran J, Gao D, Lou Y, et al. Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. J Travel Med 2020 Mar 13;27(2):taaa022. [doi: 10.1093/jtm/taaa022] [Medline: 32080723]

39. Shen X, Cai C, Li H. Quantifying socioeconomic activities and weather effects on the global spread of COVID-19 epidemic. arXiv Preprint posted online on May 20, 2020.

40. Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, et al. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. Sci Total Environ 2020 Aug 01;728:138778 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138778] [Medline: 32335405]

41. Li-Qun F, Hai-Yang Z, Han Z. Meteorological conditions and nonpharmaceutical interventions jointly determined local transmissibility of COVID-19 in 41 Chinese cities: A retrospective observational study. The Lancet Regional Health - Western Pacific 2020 Sep;2:100020-100020 [FREE Full text] [doi: 10.1016/j.lanwpc.2020.100020]

42. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020 Sep;585(7825):410-413. [doi: 10.1038/s41586-020-2293-x] [Medline: 32365354]

Abbreviations

ILI: influenza-like illness
SHAP: Shapley additive explanation
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Original Paper

Relationship Between COVID-19 Information Sources and Attitudes in Battling the Pandemic Among the Malaysian Public: Cross-Sectional Survey Study

Emma Mohamad[1], PhD; Jen Sern Tham[2], PhD; Suffian Hadi Ayub[3], PhD; Mohammad Rezal Hamzah[4], PhD; Hasrul Hashim[1], PhD; Arina Anis Azlan[1], PhD

[1]Centre for Research in Media and Communication, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[2]Department of Communication, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Seri Kembangan, Selangor, Malaysia

[3]Department of Communication, School of Arts, Sunway University, Petaling Jaya, Selangor, Malaysia

[4]Centre of Excellence for Social Innovations and Sustainability, Faculty of Applied and Human Sciences, Universiti Malaysia Perlis, Kangar, Perlis, Malaysia

Corresponding Author:
Arina Anis Azlan, PhD
Centre for Research in Media and Communication
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia
43600 UKM
Bangi, Selangor, 43600
Malaysia
Phone: 60 389215456
Email: arina@ukm.edu.my

Abstract

Background: There are multiple media platforms and various resources available for information on COVID-19. Identifying people's preferences is key to building public confidence and planning for successful national health intervention strategies.

Objective: This study examines the sources of information for COVID-19 used by the Malaysian public and identifies those that are associated with building public confidence and positive perceptions toward the Malaysian government.

Methods: A cross-sectional online survey of 4850 Malaysian residents was conducted. Participant demographics, media use, information sources, and attitudes surrounding COVID-19 were assessed. Descriptive statistics and multiple logistic regression analyses were conducted to gauge the relationship between demographics, information sources, and attitudes toward COVID-19.

Results: Malaysians primarily used television and internet news portals to access information on COVID-19. The Malaysian Ministry of Health was the most preferred source of COVID-19 information. Respondents who referred to the Ministry of Health, television, and the Malaysian National Security Council for information were more likely to believe that the country could win the battle against COVID-19 and that the government was handling the health crisis well compared to those who referred to other information sources. Those who used the World Health Organization, friends, YouTube, family, and radio as sources of information were less likely to harbor confidence and positive belief toward combating COVID-19.

Conclusions: Managing information and sustaining public confidence is important during a pandemic. Health authorities should pay considerable attention to the use of appropriate media channels and sources to allow for more effective dissemination of critical information to the public.

(J Med Internet Res 2020;22(11):e23922) doi:10.2196/23922

KEYWORDS

XSL•FO
RenderX

Introduction

The COVID-19 health crisis has introduced new challenges to effective health information delivery around the world. Past research has indicated that media hype surrounding a health crisis has caused panic and uncertainty among the public [1]. In the context of COVID-19, this is exacerbated by the misinformation and disinformation surfacing on social media. Previous studies have shown that self-efficacy plays an important role in predicting the health behaviors of the public [2]. However, in a pandemic situation, there is an increased reliance on the government and health authorities to manage the problem. The need to provide clear, consistent, and credible information surrounding the COVID-19 pandemic is key to disease mitigation and control.

One study has revealed that health information behaviors in Malaysia are highly situational and that research on health information seeking in Malaysia is still very limited [3]. Health information seeking behaviors and sources differ based on the type of disease and demographic group. Local studies on the H1N1 pandemic have found that Malaysians gathered information from the news media, social media, and authorized websites, such as the MyHealth Portal under the Ministry of Health (MOH) [2,4]. A separate study found that a majority of respondents felt that the internet was useful for assisting them in making health decisions [5].

It has been almost 1 year since COVID-19 emerged at the end of 2019. Countries all over the world are at various stages of disease mitigation. As of October 29, 2020, Malaysia has recorded 29,441 total COVID-19 cases, with 18,499 recoveries and 246 total deaths [6]. In terms of disease management, Malaysia is perceived as one of the successful countries, and this is largely due to the implementation of the Movement Control Order (MCO) on March 18, 2020 [7-9]. The country has since progressed to the third stage of the MCO called the Recovery MCO, which began on June 10, 2020. Under the Recovery MCO, Malaysians have been allowed to return to work, schools and businesses have been allowed to reopen, and travel restrictions have been loosened.

Since the implementation of the MCO, the Malaysian public has been dependent on both local and international media for information related to COVID-19. Official information and local updates on the disease were broadcast via press conferences and regular messaging through multiple telecommunication channels, including traditional media, digital media platforms, and social media. In Malaysia, these channels include government and privately owned media.

Although information is available through various local and international sources, the MOH Malaysia and the Malaysian National Security Council (NSC) are the official sources responsible for communicating information about COVID-19 and its control in the country. As such, both the MOH and NSC have made concerted efforts to maintain a consistent and reliable flow of information to the public through both traditional and new media sources. Credible spokespeople from the MOH and NSC are chosen to deliver COVID-19 updates to ensure the clarity, trustworthiness, and congruence of information. Never

before has Malaysia faced a health crisis that has required a response of this scale. Current literature emphasizes the importance of communication in a crisis for public reassurance and order [10,11]. The significant role of local media in communicating messages to the public during the MCO has also been highlighted, specifically in urging people to stay home, practice good personal hygiene, and observe social distancing when in public areas [12].

In Malaysia, the challenge of the pandemic was further complicated by the resignation of then-Prime Minister Mahathir Mohamad on February 24, 2020. The sudden change of government that occurred as the threat of COVID-19 grew introduced new uncertainties to an already volatile situation. The new government faced heightened pressure to instill confidence and trust among the people [13]. The government's ability to manage the spread of the virus and its capacity to provide quality health care for patients were important components for building that trust.

With the enforcement of the MCO, the Malaysian authorities demonstrated that they viewed the situation seriously and intended to take the necessary action to curb the spread of the virus [14]. However, implementing strict MCO rulings carries its own risks. Citizens in countries around the world have responded variably to lockdown and control measures [15]. Some have acted in retaliation to these measures, citing distrust and skepticism in health authorities [16].

Strict enforcement of the MCO by Malaysian authorities, such as the NSC and the Royal Malaysia Police, could have potentially caused backlash from the Malaysian public. However, its implementation has resulted in improved compliance with the MCO restrictions [17]. Additionally, a recent study has indicated that the Malaysian public has a positive perception on the ability of the government to control the spread of COVID-19 and high confidence that the country could win the battle against the pandemic [18].

However, this confidence and positive perception are not without challenges. This study examines where the Malaysian public source their information on COVID-19 and who they refer to for reliable information. Subsequently, this study identifies sources that are more likely to contribute to higher confidence in combating the pandemic and positive perceptions toward the government among the Malaysian public.

Methods

Participants

This cross-sectional study was conducted online between March 27 and April 3, 2020. At the time of data collection, the number of local COVID-19 cases increased from 2161 on March 27, 2020 to 3333 on April 3, 2020 (out of a 32 million total population). The survey was conducted 1 week after the highest R-naught ($R0=3.55$) was recorded, which means that 1 infected person on average will transmit the disease to 3.55 people [19].

During the survey period, Malaysia was under the MCO. Hence, it was not feasible to conduct a systematic nationwide sampling procedure. In order to overcome this challenge, the researchers

distributed an online survey using the Survey Monkey website. Survey participants were recruited by posting a 1-page recruitment message onto authors' Facebook, Twitter, Instagram, and WhatsApp groups with community leaders and social media influencers. The recruitment message contained a brief introduction on the purpose of the study, the procedures of data collection, the voluntary nature of participation, declarations of anonymity and confidentiality, and notes for filling in the questionnaire, as well as links for English and Malay language versions of the online questionnaires. Members of the public aged ≥18 years and those who resided in Malaysia were eligible to participate in this study. The target sample size was 3640, which was determined by identifying the smallest acceptable size of a demographic subgroup with a 5% margin of error and 95% confidence level. A total of 5137 surveys were initiated. Only surveys that were missing less than 10% of data were retained [20]. A total of 4850 responses were included in the final analysis.

This study was approved by the Institutional Review Board of Universiti Kebangsaan Malaysia (JEP-2020-276). The study participants were given no incentive for participation. Participants gave consent to willingly participate in the survey by clicking the "continue" button, which would direct them to complete the self-administered questionnaire.

Measures

The questionnaire was developed based on previous studies [21]. The first part of the survey questionnaire involved collecting demographic data on respondents' sex, age, ethnicity, current residing state, locality (urban or rural), occupation, and household income. The second section recorded the kinds of media and sources participants used to obtain information about COVID-19. With regard to media, participants could select any of the following: television, radio, online news portals, WhatsApp, Facebook, Twitter, Instagram, and YouTube. With regard to main sources, participants could choose the MOH, WHO, NSC, family, or friends. The choices for media and main sources were thereafter categorized into the following: mass media, social media, interpersonal sources, and authorities. Answers were coded as "yes"=1 and "no"=0. The third section measured how confident respondents were in battling the COVID-19 pandemic based on the following 2 questions: (1) "Do you have confidence that Malaysia can win the battle against COVID-19?" and (2) "The government of Malaysia is handling the COVID-19 health crisis very well." Responses to these questions were coded as "no", "disagree", and "I'm not sure" responses equaling 0 and "yes" and "agree" responses equaling 1. The instrument can be found in Multimedia Appendix 1.

Data Analysis

The collected data were analyzed using SPSS software version 26 (SPSS Inc). The proportion of participants who used each common source to obtain information about COVID-19 was presented in terms of number and percentage. To evaluate whether demographic variables (ie, sex, age, occupation, current residing state, locality, and household income) were associated with the source of information used for COVID-19, we performed a series of logistic regression tests using the forward logistic regression method. For this method, each variable is entered into the model and its value assessed based on the likelihood ratio estimates. If the variable makes a significant contribution to the model, then it is retained, and vice versa [22,23]. Logistic regression was also conducted to examine the relationships between the source of information used for COVID-19 and respondents' confidence and positive attitude toward the government in handling the health crisis. ORs, 95% CIs, and their corresponding P values are reported as indicators of the magnitude and statistical significance of associations. A P value of <.05 was considered statistically significant for all analyses.

Results

In total, the data of 4850 respondents (female: n=2808, 57.9%; male: n=2042, 42.1%) were analyzed. Most of the sample were literate and able to complete the questionnaire. Table 1 presents the descriptive statistics for respondents' demographic characteristics and the sources they used for obtaining COVID-19 information. The median age of respondents was 33.00 years (range 18-73 years). Most respondents (4026/4850, 83%) were Malay/Bumiputera, and 70% (2295/4850) of respondents resided in urban areas.

The Malaysian MOH was the most referred to source of information for the majority of study respondents (1540/4850, 95.1%) during MCO implementation in Malaysia, followed by television (3542/4850, 73%) and online news portals (3125/4850, 64%). A very small proportion of the participants obtained their information from interpersonal sources, such as friends (1235/4850, 26%) and family (1065/4850, 22%).

Tables 2-4 present the logistic regression results on the associations between demographic characteristics and each source of information. The results indicated that Malay/Bumiputera and Indian women had more than 1.7 times the odds of using television (OR 2.4, 95% CI 1.5-3.8; OR 1.7, 95% CI 1.0-2.9, respectively) as an information source compared to women of other ethnicities. Compared to men, women had 1.2 times the odds of using online news portals to seek COVID-19–related information (OR 1.2, 95% CI 1.1-1.4). In terms of social media use, the results were mixed. Women aged 30-49 years were 1.6 times more likely to use Instagram and 1.7 times more likely to use Telegram as sources of information for COVID-19 than men aged >50 years (OR 1.6, 95% CI 1.2-2.1; OR 1.7, 95% CI 1.1-2.6, respectively). The results also showed that family and friends were not the preferred source of information for COVID-19.

Table 1. Descriptive statistics for participant demographic characteristics and the sources participants used for obtaining COVID-19 information (N=4850).

Variables	Statistics
Sex, n (%)	
Male	2042 (42.1)
Female	2808 (57.9)
Age (years), median	33.00
Ethnicity, n (%)	
Malay/Bumiputera	4026 (83)
Chinese	511 (10.5)
Indian	231 (4.8)
Other[a]	82 (1.7)
Locality, n (%)	
Urban	3395 (70)
Rural	1455 (30)
Occupation, n (%)	
Public servant	2173 (44.8)
Student	1125 (23.2)
Private sector	955 (19.7)
Self-employed	267 (5.5)
Not employed	195 (4)
Retiree	96 (2)
Other[b]	32 (0.7)
Income[c] (RM/month), n (%)	
≤3000	1540 (31.8)
3001-6000	1289 (26.6)
6001-9000	832 (17.2)
9001-12,000	575 (11.9)
≥12,001	614 (12.7)
Sources of information about COVID-19, n (%)	
Ministry of Health	4614 (95.1)
Television	3542 (73)
Online news portal	3125 (64.4)
Malaysian National Security Council	3069 (63.3)
Facebook	2993 (61.7)
WhatsApp	2837 (58.5)
World Health Organization	2295 (47.3)
Twitter	1281 (26.4)
Instagram	1278 (26.4)
Friends	1235 (25.5)
YouTube	1080 (22.3)
Family	1065 (22)
Radio	854 (17.6)
Telegram	344 (7.1)

XSL•FO
RenderX

[a]Other ethnicities included Punjabi, Bugis, and Eurasian, just to name a few.

[b]Other occupations included manual labor and contract/part-time work.

[c]A currency exchange rate of RM1=US $0.24 is applicable.

Table 2. Factors related to the use of COVID-19 mass media information sources.

Variables	Television		Online news portals		Radio	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Sex - Female[a]	1.188 (1.037-1.360)	.01	1.223 (1.079-1.386)	.002	0.785 (0.672-0.918)	.002
Age (years)[b]						
18-29	0.423 (0.315-0.568)	<.001	—[c]	—	—	—
30-49	0.664 (0.512-0.859)	.002	—	—	—	—
Ethnicity[d]						
Malay/Bumiputera	2.412 (1.526-3.813)	<.001	—	—	1.598 (0.789-3.233)	.19
Chinese	1.459 (0.895-2.379)	.13	—	—	1.873 (0.897-3.909)	.10
Indian	1.707 (1.002-2.909)	.049	—	—	2.887 (1.350-6.174)	.006
Urban[e]	—	—	—	—	0.776 (0.658-0.916)	.003
Occupation[f]						
Student	1.40 (0.608-3.224)	.43	1.721 (0.833-3.558)	.14	0.445 (0.200-0.987)	.046
Not employed	0.997 (0.412-2.414)	>.99	1.501 (0.692-3.259)	.30	0.409 (0.170-0.985)	.046
Retiree	0.445 (0.173-1.147)	.094	1.404 (0.612-3.224)	.42	0.392 (0.149-1.030)	.06
Private sector	0.649 (0.284-1.486)	.31	1.515 (0.732-3.137)	.26	0.525 (0.236-1.167)	.11
Public servant	0.856 (0.376-1.948)	.71	1.302 (0.634-2.677)	.47	0.526 (0.239-1.157)	.11
Self-employed	0.543 (0.23-1.276)	.16	1.725 (0.806-3.692)	.16	0.279 (0.116-0.667)	.004
Income (RM/month)[g]						
<3000	1.032 (0.814-1.307)	.80	0.397 (0.320-0.493)	<.001	1.382 (1.044-1.830)	.02
3001-6000	1.482 (1.181-1.860)	.001	0.612 (0.492-0.762)	<.001	1.449 (1.098-1.912)	.009
6001-9000	1.405 (1.101-1.793)	.006	0.662 (0.524-0.838)	.001	1.473 (1.097-1.977)	.010
9001-12,000	1.159 (.893-1.506)	.27	0.803 (0.620-1.040)	.10	1.168 (0.842-1.620)	.35

[a]Male respondents used as a reference.

[b]Respondents aged >50 years used as a reference.

[c]Not available.

[d]Respondents who stated their ethnicity as "other" used as a reference.

[e]Respondents from rural localities used as a reference.

[f]Respondents who stated their occupation as "other" used as a reference.

[g]Respondents whose income was ≥RM12,001/month used as a reference. A currency exchange rate of RM1=US $0.24 is applicable.

Table 3. Factors related to the use of COVID-19 social media information sources.

Variables	WhatsApp		Twitter		Instagram		YouTube		Telegram	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Sex - Female[a]	0.829 (0.733-0.937)	.003	__[b]	—	1.477 (1.288-1.693)	<.001	0.628 (0.548-0.720)	<.001	1.492 (1.175-1.895)	.001
Age (years)[c]										
18-29	0.388 (0.302-0.497)	<.001	6.264 (4.278-9.171)	<.001	4.194 (3.204-5.490)	<.001	0.687 (0.556-0.850)	<.001	0.990 (0.603-1.625)	.97
30-49	0.540 (0.429-0.680)	<.001	2.096 (1.474-2.982)	<.001	1.574 (1.195-2.074)	.001	0.590 (0.477-0.730)	<.001	1.653 (1.065-2.566)	.03
Ethnicity[d]										
Malay/Bumiputera	1.656 (1.053-2.604)	.03	1.198 (.713-2.011)	.50	1.426 (.827-2.457)	.20	—	—	0.730 (0.329-1.621)	.44
Chinese	0.910 (0.562-1.473)	.70	0.228 (0.128-0.407)	<.001	0.729 (0.406-1.311)	.29	—	—	0.265 (0.102-0.685)	.006
Indian	1.654 (0.980-2.789)	.06	0.246 (0.126-0.478)	<.001	1.344 (0.726-2.490)	.35	—	—	0.508 (0.188-1.374)	.18
Urban locality[e]	—	—	1.182 (1.010-1.382)	.04	—		—	—	—	—
Occupation[f]										
Student	0.672 (0.310-1.460)	.32	1.018 (0.455-2.278)	.97	—		—	—	0.516 (0.169-1.579)	.25
Not employed	0.428 (0.190-0.964)	.04	0.616 (0.262-1.445)	.27	—		—	—	0.537 (0.164-1.764)	.31
Retiree	0.795 (0.320-1.980)	.62	0.262 (0.076-0.897)	.03	—		—	—	0.110 (0.011-1.060)	.06
Private sector	0.581 (0.269-1.256)	.17	0.726 (0.326-1.617)	.43	—		—	—	0.687 (0.232-2.035)	.50
Public servant	0.783 (0.365-1.684)	.53	0.351 (0.159-0.777)	.01	—		—	—	0.601 (0.205-1.758)	.35
Self-employed	0.449 (0.202-0.997)	.049	0.382 (0.162-0.901)	.03	—		—	—	0.244 (0.070-0.854)	.03
Income (RM/month)[g]										
<3000	—	—	0.537 (0.411-0.700)	<.001	—		—	—	—	—

Variables	WhatsApp		Twitter		Instagram		YouTube		Telegram	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
3001-6000	—	—	0.736 (0.571-0.948)	.02	—	—	—	—	—	—
6001-9000	—	—	0.897 (0.684-1.175)	.43	—	—	—	—	—	—
9001-12,000	—	—	0.875 (0.651-1.176)	.38	—	—	—	—	—	—

[a]Male respondents used as a reference.

[b]Not available.

[c]Respondents aged >50 years used as a reference.

[d]Respondents who stated their ethnicity as "other" used as a reference.

[e]Respondents from rural localities used as a reference.

[f]Respondents who stated their occupation as "other" used as a reference.

[g]Respondents whose income was ≥RM12,001/month used as a reference. A currency exchange rate of RM1=US $0.24 is applicable.

Table 4. Factors related to the use of COVID-19 interpersonal and health authority information sources.

Variables	MOH[a]		NSC[b]		Friends		Family	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Sex - Female[c]	—[d]	—	1.141 (1.006-1.296)	.04	—	—	—	—
Age (years)[e]								
18-29	—	—	0.523 (0.401-0.681)	<.001	0.500 (0.408-0.612)	<.001	0.976 (0.785-1.213)	.83
30-49	—	—	0.950 (0.756-1.194)	.66	0.590 (0.483-0.720)	<.001	0.652 (0.522-0.813)	<.001
Ethnicity[f]								
Malay/Bumiputera	3.521 (1.776-6.981)	<.001	1.619 (1.029-2.548)	.04	—	—	—	—
Chinese	1.113 (0.540-2.291)	.77	0.660 (0.407-1.069)	.09	—	—	—	—
Indian	1.465 (0.653-3.288)	.35	0.852 (0.507-1.433)	.55	—	—	—	—
Urban locality[g]	—		—		—	—	—	—
Occupation[h]								
Student	—	—	1.227 (.576-2.616)	.60	—	—	—	—
Not employed	—	—	1.001 (0.450-2.223)	>.99	—	—	—	—
Retiree	—	—	0.826 (0.347-1.966)	.67	—	—	—	—
Private sector	—	—	1.179 (0.556-2.501)	.67	—	—	—	—
Public servant	—	—	1.080 (0.513-2.274)	.84	—	—	—	—
Self-employed	—	—	0.687 (0.315-1.496)	.34	—	—	—	—
Income (RM/month)[i]								
<3000	1.059 (0.716-1.567)	.77	0.783 (0.624-0.982)	.03	—	—	—	—
3001-6000	1.630 (1.060-2.506)	.03	0.962 (0.776-1.194)	.73	—	—	—	—
6001-9000	1.876 (1.142-3.080)	.01	1.056 (0.837-1.331)	.65	—	—	—	—
9001-12,000	1.465 (0.871-2.462)	.15	1.034 (0.804-1.331)	.79	—	—	—	—

[a]MOH: Ministry of Health.

[b]NSC: National Security Council.

[c]Male respondents used as a reference.

[d]Not available.

[e]Respondents aged >50 years used as a reference.

[f]Respondents who stated their ethnicity as "other" used as a reference.

[g]Respondents from rural localities used as a reference.

[h]Respondents who stated their occupation as "other" used as a reference.

[i]Respondents whose income was ≥RM12,001/month used as a reference. A currency exchange rate of RM1=US $0.24 is applicable.

Table 5 focuses on the relationship between the sources of information used by respondents and their confidence in battling the COVID-19 pandemic and positive attitudes toward the Malaysian government in handling the COVID-19 crisis. The results showed that using the MOH, television, and the Malaysian NSC as information sources were significantly associated with confidence and positive attitudes among the respondents. Specifically, people who used the MOH, NSC, and television as sources of information for COVID-19 had more than 1.7 times the odds of having confidence in battling the pandemic (OR 2.8, 95% CI 1.7-4.7; $P<.001$;OR 1.9, 95% CI 1.4-2.6; $P<.001$; OR 1.8, 95% CI 1.3-2.5; $P=.001$, respectively) and positive attitudes toward the Malaysian government (OR 2.3, 95% CI 1.6-3.3; $P<.001$; OR 2.0, 95% CI 1.6-2.4; $P<.001$; OR 1.8, 95% CI 1.4-2.2; $P<.001$, respectively) compared to those who did not use these information sources.

They also agreed that the Malaysian government handled the COVID-19 crisis very well. Respondents who used Instagram as their source of information were more likely to believe that the crisis was handled well by the government than those who did not (OR 1.4, 95% CI 1.1-1.8; $P=.007$), but this was not significantly associated with their confidence in battling the COVID-19 pandemic. Interestingly, people who sourced information from the World Health Organization (WHO) were less likely to harbor confidence and positive belief toward the government in handling the COVID-19 crisis (OR 0.44, 95% CI 0.3-0.6; OR 0.63, 95% CI 0.5-0.8, respectively). With regard to online news portals, WhatsApp, Twitter, and Telegram, a nonsignificant result was noted between each information source used by a respondent and confidence in battling the COVID-19 pandemic.

Table 5. Predicting general confidence and positive attitude toward the Malaysian government based on respondents' sources of COVID-19–related information.

Variable (Use[a]=1)	Do you have confidence that Malaysia can win the battle against COVID-19? (Yes=1)		The government of Malaysia is handling the COVID-19 health crisis very well. (Agree=1)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
MOH[b]	2.828 (1.714-4.667)	<.001	2.313 (1.625-3.294)	<.001
Television	1.765 (1.262-2.468)	.001	1.778 (1.430-2.210)	<.001
Online news portal	__[c]	—	—	—
NSC[d]	1.895 (1.360-2.641)	<.001	1.965 (1.600-2.415)	<.001
WhatsApp	—	—	—	—
WHO[e]	0.436 (0.312-0.609)	<.001	0.631 (0.514-0.774)	<.001
Twitter	—	—	—	—
Instagram	—	—	1.401 (1.095-1.791)	.007
Friends	0.691 (0.487-0.979)	.04	—	—
YouTube	—	—	0.687 (0.542-0.871)	.002
Family	—	—	0.734 (0.581-0.928)	.01
Radio	—	—	0.729 (0.557-0.955)	.02
Telegram	—	—	—	—

[a]Respondents who did not use the information sources were used as a reference.

[b]MOH: Ministry of Health.

[c]Not available.

[d]NSC: National Security Council.

[e]WHO: World Health Organization.

Discussion

Principal Findings

The findings of this study indicate that during the MCO, Malaysian people mainly relied on television for updated information related to COVID-19, followed by internet news portals. Apart from those, we also found that respondents from certain demographic groups favored other media platforms. In particular, young adults aged 18-29 years preferred referring to the WHO and Instagram for information; Malays preferred referring to the MOH, television, NSC, and WhatsApp; and

young urbanites preferred referring to the WHO and Twitter. Variety in the use of information sources by demographic groups was also found in a previous study conducted in Taiwan [24].

In terms of the sources of COVID-19–related information used by our respondents, 95.1% (4614/4850) of the respondents referred to the MOH and 63.3% (3069/4850) referred to the NSC. This is because the MOH is the local authority on public health and the primary source of information on COVID-19. However, the WHO, the international governing body, was not favored by the majority of Malaysians (2295/4850, 47.3%). Possible reasons for the low preference toward using the WHO

as an information source is the language barrier and a higher reliance on local health authorities among the Malaysian public.

During the MCO period, Malaysians used television as their main source of information. On television, COVID-19–related information from the MOH was consistently and constantly updated by way of daily briefings and press conferences. The public was reassured that proper actions were being taken through messages delivered by strong spokespeople, public service announcements, the use of simple language, and the use of visuals, such as informational graphics and short videos.

This study also indicated that confidence in winning the battle against the COVID-19 pandemic was more than 1.7 times higher in those who received information from the MOH, NSC, and television. Public health information has been successfully communicated using television in times of a health crisis, during which much needed knowledge for disease prevention and control can strengthen public confidence in the ability of the government to manage the pandemic [25,26].

Although the internet and social media are popular tools for information searching, this study found that the use of social media was less likely to stimulate public confidence toward government efforts in handling the COVID-19 pandemic. This likely means that social media use during the MCO was for social engagement and entertainment purposes. Our findings offer a different perspective from past studies, which found that social media, such as Facebook and Twitter, served as useful tools for raising awareness and promoting intervention measures for the H1N1 pandemic [27-32].

In conclusion, despite the abundance of media sources, traditional media was proven to be the most preferred platform for obtaining information during the COVID-19 health crisis [33]. Local authorities were also seen as champions in the dissemination of information. These findings suggest that the best way to communicate to the public in times of a health crisis is via local authorities and traditional mass media outlets.

Limitations

This study had a few limitations, especially in the recruitment of respondents. First, the dissemination of the questionnaires was done via social media, thereby limiting the sample to those with internet connections. This may have excluded marginalized groups with limited or no internet access. Second, internet and social media users may not be a true representation of the population, as internet and social media sources are prone to bias in demographics, such as age, sex, location, and income category [34]. Compared to the Malaysian population, our study sample was somewhat skewed. According to the Department of Statistics Malaysia, the percentages of males and females in Malaysia are 51.5% and 48.5%, respectively. Our sample of 4850 respondents consisted of 2042 (42.1%) males and 2808 (57.9%) females. In terms of urban and rural distribution, 77% of Malaysians live in urban areas and 23% live in rural areas.

In our study, 70% (n=3395) of respondents lived in urban areas and 30% (n=1455) lived in rural areas. Additionally, the median income for Malaysia in 2019 was RM5873 (US $1418.09) per month. In this study, 58.4% (n=2829) of respondents earned a household income of less than RM6000 (US $1448.75) per month. In terms of racial distribution in Malaysia, 69.7% of people are Malay/Bumiputera, 22.6% are Chinese, and 6.9% are Indian. In our study, 83% (n=4026) of respondents were Malay/Bumiputera, 10.5% (n=511) were Chinese, and 4.8% (n=82) were Indian [35]. In terms of employment, only 11% of Malaysians work in the public sector. However, in this study, 44.8% (n=2173) of our respondents worked in the public sector [36]. Third, the questionnaire was distributed during the first phase of the COVID-19 MCO in Malaysia. Therefore, it only illustrates the perspectives of the Malaysian public during that time. Lastly, the conclusion of this study reflects information sources and public confidence in the context of Malaysia and may not be applicable to other populations.

Conclusions

Managing information during a pandemic is vehemently important to ensure that public health remains the top priority in the government agenda. Similarly, the public too must be equipped with accurate and timely information to keep abreast of COVID-19 news and the appropriate preventive strategies. As new scientific discoveries are made and new information becomes available, the government and relevant authorities must be able to react quickly. This is especially important as the world confronts the proliferation of misinformation, fake news, and conspiracy theories surrounding COVID-19. Old information must be replaced with new information, erroneous facts must be corrected, and public announcements and instructions must be promptly disseminated [37].

The results of this study contribute to the understanding of sources used to obtain COVID-19 information and their relationships with building public confidence in the face of a pandemic. The right tools and channels are key determinants for ensuring effective public health information delivery. The ability of health authorities to understand and utilize these tools plays an important role in building confidence among the intended public.

Additionally, the findings of this study can aid public health educators in the strategic use of information platforms and sources to effectively communicate COVID-19–related information. The results are also useful for developing strategic communication plans to cope with the increasing spread of misinformation and disinformation surrounding COVID-19 in Malaysia.

Future research is needed to understand the changing use of communication tools as the pandemic evolves. Its association and impact on public perception should also be evaluated at different stages of disease mitigation to observe any notable changes.

Acknowledgments

The authors are grateful to the Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, particularly Associate Professor Dr. Kadaruddin Aiyub and Noraznita Anor Basri, for their assistance. We would also like to express our sincere thanks to Dr. Hasidah Mohd Sidek for her editorial contribution and to all members of the Malaysian public who participated in the survey.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey Questions.
[DOCX File, 20KB - jmir_v22i11e23922_app1.docx]

References

1. Klemm C, Das E, Hartmann T. Swine flu and hype: a systematic review of media dramatization of the H1N1 influenza pandemic. J Risk Res 2014 Jun 20;19(1):1-20 [FREE Full text] [doi: 10.1080/13669877.2014.923029]
2. Lin L, Savoia E, Agboola F, Viswanath K. What have we learned about communication inequalities during the H1N1 pandemic: a systematic review of the literature. BMC Public Health 2014 May 21;14:484 [FREE Full text] [doi: 10.1186/1471-2458-14-484] [Medline: 24884634]
3. Hamzah MR, Mohammad EMW, Abdullah MY, Ayub SH. Scenario on health information seeking in Malaysia: a systematic review. Journal of Human Development and Communication. 2015. URL: https://johdec.unimap.edu.my/images/PDF/Vol_4_2015/Vol_4_2015_7-20.pdf [accessed 2020-10-16]
4. Tobi SNM, Masrom M, Rahaman SASA. Healthcare Portals and the Usage Trends in Malaysia. Adv Sci Lett 2017 Apr 01;23(4):2857-2860. [doi: 10.1166/asl.2017.7722]
5. Wong LP. Premenstrual syndrome and dysmenorrhea: urban-rural and multiethnic differences in perception, impacts, and treatment seeking. J Pediatr Adolesc Gynecol 2011 Oct;24(5):272-277. [doi: 10.1016/j.jpag.2011.03.009] [Medline: 21600811]
6. Southeast Asia Covid-19 tracker. Center For Strategic & International Studies. URL: https://www.csis.org/programs/southeast-asia-program/southeast-asia-covid-19-tracker-0 [accessed 2020-10-29]
7. Malaysia recognised as one of the most successful in handling COVID-19. Prime Minister Office of Malaysia Official Website. 2020 Jul 06. URL: https://www.pmo.gov.my/2020/07/malaysia-recognised-as-one-of-most-successful-in-handling-covid-19/ [accessed 2020-07-16]
8. Nixon S. Commentary: Malaysia is beating all these brutal COVID-19 expectations. Channel News Asia. 2020 May 04. URL: https://www.channelnewsasia.com/news/commentary/malaysia-beating-brutal-covid-19-coronavirus-infections-deaths-12691586 [accessed 2020-07-16]
9. Omar ARC, Ishak S, Jusoh MA. The impact of Covid-19 Movement Control Order on SMEs' businesses and survival strategies. Malaysian Journal of Society and Space. 2020 May 30. URL: https://www.researchgate.net/publication/341765167 [accessed 2020-11-04]
10. Mohamad E, Azlan AA. COVID-19 and communication planning for health emergencies. Malaysian Journal of Coummunication. 2020. URL: http://ejournals.ukm.my/mjc/article/view/38718 [accessed 2020-07-22]
11. Sannusi SN, Siarap K. The role of public relations and communication crisis: a study on handling of SARS outbreak by ministry of health Malaysia. e-Bangi Journal of Social Sciences and Humanities. 2014. URL: http://ejournals.ukm.my/ebangi/article/view/11095/3606 [accessed 2020-11-04]
12. Tang KHD. Movement control as an effective measure against Covid-19 spread in Malaysia: an overview. Z Gesundh Wiss 2020 Jun 13:1-4 [FREE Full text] [doi: 10.1007/s10389-020-01316-w] [Medline: 32837842]
13. Review Procedure, Way of Govt Administration, Post COVID-19 – Muhyiddin. Prime Minister's Office of Malaysia Official Website. 2020 Jul 06. URL: https://www.pmo.gov.my/2020/07/review-procedure-way-of-govt-administration-post-covid-19-muhyiddin/ [accessed 2020-07-16]
14. Azman H. Cognitive, effective and conative message strategies in official announcements to effects positive response for the movement control order to suppress the Covid-19 pandemic. The Southeast Asian Journal of English Language Studies. 2020 Jun. URL: https://www.researchgate.net/publication/342548970_COVID-19_INSIGHTS_AND_LINGUISTIC_METHODS [accessed 2020-11-04]
15. Stella M, Restocchi V, De Deyne S. #lockdown: Network-Enhanced Emotional Profiling in the Time of COVID-19. Big Data and Cognitive Computing. 2020 Jun 16. URL: https://www.mdpi.com/2504-2289/4/2/14/htm [accessed 2020-11-04]
16. Ling G. Effects of the coronavirus (COVID-19) pandemic on social behaviours: from a social dilemma perspective. Technium Social Sciences Journal. 2020. URL: https://ideas.repec.org/a/tec/journl/v7y2020i1p312-320.html [accessed 2020-07-16]

XSL•FO
RenderX

17. Zolkepli F, Sivanandam H. MCO compliance at 97% but 482 arrests over March 26?27. The Star. 2020 Mar 28. URL: https://www.thestar.com.my/news/nation/2020/03/28/mco-compliance-at-97-but-482-arrests-over-march-26-27 [accessed 2020-07-20]

18. Azlan A, Hamzah M, Sern TJ, Ayub S, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: A cross-sectional study in Malaysia. PLoS One 2020 May 21;15(5):e0233668. [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]

19. Arif MM. COVID-19 Weekly Round-up: 287 New Cases, Malaysia May Face Third Wave. Bernama. 2020 Oct 02. URL: https://www.bernama.com/en/features/news.php?id=1885974 [accessed 2020-10-18]

20. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis, Seventh Edition, Pearson New International Edition. Essex: Pearson Education Limited; 2014.

21. Zhong BL, Luo W, Li HM, Zhang QQ, Liu XG, Li WT, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]

22. Chua YP. Statistik Penyelidikan Lanjutan (Buku 5). Kuala Lumpur: McGraw-Hill (M); 2009.

23. Field A. Discovering Statistics Using IBM SPSS Statistics, 5th Edition. London: Sage Publications Ltd (UK); 2018.

24. Wang PW, Lu WH, Ko NY, Chen YL, Li DJ, Chang YP, et al. COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]

25. Smith GD, Ng F, Li WHC. COVID-19: Emerging compassion, courage and resilience in the face of misinformation and adversity. J Clin Nurs 2020 May;29(9-10):1425-1428 [FREE Full text] [doi: 10.1111/jocn.15231] [Medline: 32155302]

26. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020 Apr 01;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]

27. Frew PM, Painter JE, Hixson B, Kulb C, Moore K, del Rio C, et al. Factors mediating seasonal and influenza A (H1N1) vaccine acceptance among ethnically diverse populations in the urban south. Vaccine 2012 Jun 13;30(28):4200-4208 [FREE Full text] [doi: 10.1016/j.vaccine.2012.04.053] [Medline: 22537991]

28. Reynolds BJ. Building trust through social media. CDC's experience during the H1N1 influenza response. Mark Health Serv 2010;30(2):18-21. [Medline: 20550001]

29. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]

30. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199. [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]

31. Ding H, Zhang J. Social Media and Participatory Risk Communication during the H1N1 Flu Epidemic: A Comparative Study of the United States and China. China Media Research. 2010 Jan. URL: http://chatrlab.ucdavis.edu/uploads/8/4/7/8/84783462/social_media_and_participatory_risk_communication_during_the_h1n1_flu_epidemic_-_a_comparative_study.pdf [accessed 2020-11-04]

32. Liu BF, Kim S. How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for U.S. health communicators. Public Relat Rev 2011 Sep;37(3):233-244. [doi: 10.1016/j.pubrev.2011.03.005]

33. Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020 Mar 25;368:m1160. [doi: 10.1136/bmj.m1160] [Medline: 32213480]

34. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

35. Current Population Estimates, Malaysia, 2020. Department of Statistics Malaysia Official Portal. URL: https://tinyurl.com/y5tjjbkh [accessed 2020-10-26]

36. Annual Report 2019. Public Service Department Malaysia Website. URL: https://docs.jpa.gov.my/docs/pnerbitan/ltahunan/flipbook/lp2019/22/#zoom=z [accessed 2020-10-26]

37. Okan O, Sørensen K, Messer M. COVID-19: A guide to good practice on keeping people well informed. The Conversation. 2020 Mar 19. URL: https://theconversation.com/covid-19-a-guide-to-good-practice-on-keeping-people-well-informed-134046 [accessed 2020-07-20]

Abbreviations

MCO: Movement Control Order
MOH: Ministry of Health
NSC: National Security Council
WHO: World Health Organization

XSL•FO
RenderX

<u>Original Paper</u>

Health Literacy, eHealth Literacy, Adherence to Infection Prevention and Control Procedures, Lifestyle Changes, and Suspected COVID-19 Symptoms Among Health Care Workers During Lockdown: Online Survey

Binh N Do[1,2,3*], MD, PhD; Tien V Tran[1,2,3*], MD, PhD; Dung T Phan[4,5], RN, MPH, PhD; Hoang C Nguyen[6,7], MD, PhD; Thao T P Nguyen[8,9], MD; Huu C Nguyen[10,11], MD, PhD; Tung H Ha[12], MD, PhD; Hung K Dao[13], MD, PhD; Manh V Trinh[14], MD; Thinh V Do[15], MD; Hung Q Nguyen[16], MD, MSc; Tam T Vo[17], MD; Nhan P T Nguyen[18], MD, PhD; Cuong Q Tran[19,20], MD, MPH, MBA; Khanh V Tran[21], MD; Trang T Duong[22], RN, MPH; Hai X Pham[23], MD; Lam V Nguyen[24,25], MD, PhD; Kien T Nguyen[26], MA; Peter W S Chang[27,28], MD, ScD; Tuyen Van Duong[29], BNSc, MNSc, PhD

[1]Department of Infectious Diseases, Vietnam Military Medical University, Hanoi, Vietnam

[2]Division of Military Science, Military Hospital 103, Hanoi, Vietnam

[3]Director Office, Military Hospital 103, Hanoi, Vietnam

[4]Faculty of Nursing, Hanoi University of Business and Technology, Hanoi, Vietnam

[5]Nursing Office, Viet Duc University Hospital, Hanoi, Vietnam

[6]Director Office, Thai Nguyen National Hospital, Thai Nguyen, Vietnam

[7]President Office, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam

[8]Health Management Training Institute, Hue University of Medicine and Pharmacy, Thua Thien Hue, Vietnam

[9]Department of Health Economics, Corvinus University of Budapest, Budapest, Hungary

[10]Department of Thoracic and Cardiovascular Surgery, E Hospital, Hanoi, Vietnam

[11]Director Office, E Hospital, Hanoi, Vietnam

[12]Director Office, General Hospital of Agricultural, Hanoi, Vietnam

[13]Director Office, Bac Ninh Obstetrics and Pediatrics Hospital, Bac Ninh, Vietnam

[14]Director Office, Quang Ninh General Hospital, Quang Ninh, Vietnam

[15]Director Office, Bai Chay Hospital, Quang Ninh, Vietnam

[16]Director Office, Quang Ninh Obstetrics and Pediatrics Hospital, Quang Ninh, Vietnam

[17]Director Office, Trieu Phong District Health Center, Quang Tri, Vietnam

[18]General Planning Department, Da Nang Oncology Hospital, Da Nang, Vietnam

[19]Director Office, Thu Duc District Health Center, Ho Chi Minh, Vietnam

[20]Faculty of Health, Mekong University, Vinh Long, Vietnam

[21]Director Office, Hospital District 2, Ho Chi Minh, Vietnam

[22]Nursing Office, Tan Phu District Hospital, Ho Chi Minh, Vietnam

[23]Director Office, District 9 Health Center, Ho Chi Minh, Vietnam

[24]President Office, Can Tho University of Medicine and Pharmacy, Can Tho, Vietnam

[25]Aesthetic Plastic Surgery & Skin Care Center, Can Tho University of Medicine and Pharmacy Hospital, Can Tho, Vietnam

[26]Department of Health Education, Faculty of Social Sciences, Behavior and Health Education, Hanoi University of Public Health, Hanoi, Vietnam

[27]School of Medicine, Chung Shan Medical University, Taichung, Taiwan

[28]Department of Public Health and Community Medicine, Tufts University School Of Medicine, Boston, MA, United States

[29]School of Nutrition and Health Sciences, Taipei Medical University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Tuyen Van Duong, BNSc, MNSc, PhD
School of Nutrition and Health Sciences
Taipei Medical University
250 Wuxing St
Taipei, 11031

Taiwan
Phone: 886 2 2736 1661 ext 6545
Email: tvduong@tmu.edu.tw

Abstract

Background: The COVID-19 pandemic has imposed a heavy burden on health care systems and governments. Health literacy (HL) and eHealth literacy (as measured by the eHealth Literacy Scale [eHEALS]) are recognized as strategic public health elements but they have been underestimated during the pandemic. HL, eHEALS score, practices, lifestyles, and the health status of health care workers (HCWs) play crucial roles in containing the COVID-19 pandemic.

Objective: The aim of this study is to evaluate the psychometric properties of the eHEALS and examine associations of HL and eHEALS scores with adherence to infection prevention and control (IPC) procedures, lifestyle changes, and suspected COVID-19 symptoms among HCWs during lockdown.

Methods: We conducted an online survey of 5209 HCWs from 15 hospitals and health centers across Vietnam from April 6 to April 19, 2020. Participants answered questions related to sociodemographics, HL, eHEALS, adherence to IPC procedures, behavior changes in eating, smoking, drinking, and physical activity, and suspected COVID-19 symptoms. Principal component analysis, correlation analysis, and bivariate and multivariate linear and logistic regression models were used to validate the eHEALS and examine associations.

Results: The eHEALS had a satisfactory construct validity with 8 items highly loaded on one component, with factor loadings ranked from 0.78 to 0.92 explaining 76.34% of variance; satisfactory criterion validity as correlated with HL (ρ=0.42); satisfactory convergent validity with high item-scale correlations (ρ=0.80-0.84); and high internal consistency (Cronbach α=.95). HL and eHEALS scores were significantly higher in men (unstandardized coefficient [B]=1.01, 95% CI 0.57-1.45, P<.001; B=0.72, 95% CI 0.43-1.00, P<.001), those with a better ability to pay for medication (B=1.65, 95% CI 1.25-2.05, P<.001; B=0.60, 95% CI 0.34-0.86, P<.001), doctors (B=1.29, 95% CI 0.73-1.84, P<.001; B 0.56, 95% CI 0.20-0.93, P=.003), and those with epidemic containment experience (B=1.96, 95% CI 1.56-2.37, P<.001; B=0.64, 95% CI 0.38-0.91, P<.001), as compared to their counterparts, respectively. HCWs with higher HL or eHEALS scores had better adherence to IPC procedures (B=0.13, 95% CI 0.10-0.15, P<.001; B=0.22, 95% CI 0.19-0.26, P<.001), had a higher likelihood of healthy eating (odds ratio [OR] 1.04, 95% CI 1.01-1.06, P=.001; OR 1.04, 95% CI 1.02-1.07, P=.002), were more physically active (OR 1.03, 95% CI 1.02-1.03, P<.001; OR 1.04, 95% CI 1.03-1.05, P<.001), and had a lower likelihood of suspected COVID-19 symptoms (OR 0.97, 95% CI 0.96-0.98, P<.001; OR 0.96, 95% CI 0.95-0.98, P<.001), respectively.

Conclusions: The eHEALS is a valid and reliable survey tool. Gender, ability to pay for medication, profession, and epidemic containment experience were independent predictors of HL and eHEALS scores. HCWs with higher HL or eHEALS scores had better adherence to IPC procedures, healthier lifestyles, and a lower likelihood of suspected COVID-19 symptoms. Efforts to improve HCWs' HL and eHEALS scores can help to contain the COVID-19 pandemic and minimize its consequences.

Introduction

COVID-19, the disease caused by SARS-CoV-2, has created unprecedented challenges worldwide [1-6], with significant socioeconomic burdens [7], morbidity, and mortality [8,9]. Multidisciplinary and multidimensional approaches are required to contain the pandemic [10-12]. Social and behavioral changes [13,14] and improving health literacy (HL) [15] and eHealth literacy (as measured by the eHealth Literacy Scale [eHEALS]) [16] are highly recommended to control this global health crisis while effective treatments and vaccines are unavailable.

Lockdown measures were applied in many countries, including Vietnam [17]. Lockdown is a necessary public health approach to contain COVID-19. During the lockdown period, health care workers (HCWs) still needed to work and provide health services to patients, which made them more vulnerable to

COVID-19 infection. In addition, the pandemic itself and social distancing measures have had negative impacts on psychological health [18,19] and health-related behavior (eg, unhealthy eating habits and behaviors, less physical activity, and more smoking [20-23] and drinking to cope amid the COVID-19 pandemic [24]).

Online consultations from hospitals and health care centers were found to be a safe and effective way to reduce the negative effects of the pandemic [25]. In addition, social media can provide insights for effective communication among public health authorities [26], researchers [27], and the public [28-30]. However, social media spreads panic and anxiety among the public [31]. Disinformation and misinformation have been raised as highly concerning issues for the public [32-34]. HL plays an important role in evaluating online health information [35]. HL is critical for people who interact with the digital world, with

its diverse information and sources [36], especially during the COVID-19 pandemic. HCWs play an important role in supporting the public to combat misinformation and disinformation [28,37]. The HCWs' consultations cannot be altered by social media networks [37,38]. Continuous training and education have been recognized as effective approaches to improving HCWs' HL, further improving health care delivery [39,40], communication [39,41,42], shared decision making [43], and patient health outcomes [40]. Furthermore, raising awareness of behavioral pitfalls could support appropriate behavioral changes and containment of the crisis [44]. HL and eHealth literacy are more important than ever due to the COVID-19 pandemic [16,45]. However, these issues have been underestimated during the pandemic [15].

We evaluated the psychometric properties of the eHEALS and examined the predictors of HL and eHEALS scores. We also examined the associations between HL and eHEALS scores with adherence to infection prevention and control (IPC) measures, lifestyle changes, and suspected COVID-19 symptoms among HCWs during the lockdown period in Vietnam.

Methods

Study Design and Settings

A cross-sectional study was conducted with HCWs April 6-19, 2020, using online-based questionnaires (Text 1 in Multimedia Appendix 1). The HCWs were recruited from 12 hospitals and 3 health centers across Vietnam, including 8 hospitals in the nation's north, 1 hospital and 1 health center in the central region, and 3 hospitals and 2 health centers in the country's south.

Study Participants and Data Collection

No HCWs (doctors and nurses) in our study had provided any direct care or had contact with patients with COVID-19. A total sample of 5209 HCWs (out of 11,517 possible participants) completed an online survey. The studied and possible participants from public hospitals and health centers are presented in Table 1.

Table 1. Participants from the studied hospitals and health centers by geographic location.

Geographic location and hospital/health center	Possible participants	Studied participants
North		
Ha Noi city		
Military Hospital 103	1660	177
E hospital	1125	335
General Hospital of Agricultural	555	424
Thai Nguyen province		
Thai Nguyen National Hospital	1186	988
Bac Ninh city		
Bac Ninh Obstetrics and Pediatrics Hospital	391	364
Quang Ninh province		
Quang Ninh General Hospital	922	675
Bai Chay Hospital	819	476
Quang Ninh Obstetrics and Pediatrics Hospital	478	290
Center		
Quang Tri province		
Trieu Phong District Health Center	271	203
Da Nang city		
Da Nang Oncology Hospital	555	134
South		
Ho Chi Minh city		
Tan Phu District Hospital	530	241
Hospital District 2	812	318
District 9 Health Center	170	102
Thu Duc District Health Center	302	291
Can Tho city		
Can Tho University of Medicine and Pharmacy Hospital	424	191
Total	11,517	5209

Vietnam applied a nationwide lockdown measure April 1-15, 2020 [17], which was extended to April 22, 2020 [46,47]. In this study, HCWs took the online survey during the lockdown period. We used Google Forms to design and conduct the survey. The online survey links were sent to HCWs by researchers via email, Messenger, or Zalo. QR codes were also displayed in different departments of hospitals and health centers. It took about 15 minutes to complete the questionnaire. All survey questions were mandatory; therefore, there is no missing data in our study. All responses were exported to Google Sheets and saved on Google Drive. Finally, the data was coded, cleaned, and analyzed confidentially by researchers.

Measures

Sociodemographics

HCWs reported their age (21-40 years versus 41-60 years), gender (woman versus man), marital status (never married versus ever married), ability to pay for medication (very or fairly difficult versus very or fairly easy), social status (low versus middle to high), profession (doctor, nurse, or other, the last of which included medical technicians, midwives, pharmacists, pharmacy technicians, administrative staff, catering staff, and cleaners), type of health care facility (second-line versus frontline, the latter of which includes the outpatient department, emergency department, quarantine and isolation areas, medical imaging and laboratory diagnosis department, and patient administration areas), and previous epidemic (eg, SARS, tuberculosis, influenza A) containment experience (no versus yes). Additionally, comorbidity was assessed using the Charlson comorbidity index items [48,49].

Health Literacy

A 12-item short-form health literacy questionnaire (HLS-SF12) was used. The questionnaire was validated and used in Asian countries [50], including Vietnam [51-54]. HCWs rated their perceived difficulty of items based on a 4-point Likert scale from 1=Very difficult to 4=Very easy. The health literacy index

score was standardized to an unified metric from 0 to 50, with higher scores representing better health literacy [50,55].

eHealth Literacy

The widely used eHealth literacy scale (eHEALS) with 8 items was used to assess HCWs' eHealth literacy skills [56]. The questionnaire was translated into Vietnamese by the researchers. The content was then validated by an expert panel (28 medical doctors, 7 nurses, and 9 nutrition and public health professionals). The expert panel suggested retaining the original rating and scoring system. HCWs rated their experiences using the internet for health information based on a 5-point Likert scale from 1=Strongly disagree to 5=Strongly agree. Scores range from 8 to 40, with high eHEALS scores representing better eHealth literacy.

Adherence to IPC Procedures

Participants reported the practices and activities performed related to COVID-19 IPC during health care interactions. The questionnaire was adapted from the interim guidance of the World Health Organization [57]. The IPC measures included the following: (1) wearing single-use gloves, (2) wearing a medical mask, (3) wearing a face shield or goggles/protective glasses, (4) wearing a disposable gown, (5) removing and replacing personal protective equipment (PPE) according to protocol, (6) performing hand hygiene before and after touching patients, (7) performing hand hygiene before and after performing any clean or aseptic procedure, (8) performing hand hygiene after exposure to body fluid, (9) performing hand hygiene after touching patients' surroundings, and (10) decontaminating high-touch surfaces. HCWs quantified the frequency of performance of the above IPC procedures as recommended, where 0=Never, 1=Rarely (<20% of the time), 2=Occasionally (20% to <50% of the time), 3=Most of the time (50% to <95% of the time), and 4=Always (≥95% of the time). The total performance score of the 10 activities ranged from 0 to 40, with higher scores indicating better adherence to IPC procedures.

Lifestyle Changes

HCWs reported their current smoking (never/stopped/less versus unchanged or more), drinking (never/stopped/less versus unchanged or more), physical activity (never/stopped/less versus unchanged or more), and eating (less healthy versus unchanged or healthier) behaviors as compared with that before the pandemic [54].

Suspected COVID-19 Symptoms

HCWs were screened for suspected COVID-19 symptoms [58], including the common symptoms of fever, cough, and dyspnea, along with the less common symptoms of myalgia, fatigue, sputum production, confusion, headache, sore throat, rhinorrhea, chest pain, hemoptysis, diarrhea, and nausea/vomiting. HCWs with any of those symptoms were classified as having symptoms indicative of COVID-19.

Ethical Consideration

Data Analysis

Psychometric Properties of the eHEALS

Construct Validity

The construct validity of the eHEALS was examined using principal component analysis (PCA). An Kaiser-Meyer Olkin (KMO) value ≥0.6 was set to measure sampling adequacy and a Bartlett Test of Sphericity value <0.05 was set to determine the suitability of the data for PCA [59]. The oblique rotation (Promax) method was used.

Convergent Validity

The Spearman correlation was used to check the correlations between the eHEALS and its eight items.

Criterion Validity

The Pearson correlation between eHEALS and HLS-SF12 was estimated to provide evidence of criterion validity [60].

Floor and Ceiling Effects

The percentages of the lowest and highest score among HCWs were calculated. Minimal percentages (<15%) were recommended to eliminate floor and ceiling effects [61].

Reliability Analysis

The internal consistency of the eHEALS was estimated using the Cronbach α test. A Cronbach α value ≥.70 was designated as satisfactory reliability [62].

Health Literacy, eHealth Literacy, and Associated Factors

The distributions of HL and eHEALS scores in different categories of studied variables were explored using a one-way analysis of variance (ANOVA) test. In addition, bivariate and multivariate linear regression models were used to examine predictors of HL and eHEALS scores and to investigate the associations of HL and eHEALS scores with adherence to IPC measures. Next, bivariate and multivariate logistic regression models were used to examine the associations of HL and eHEALS scores with lifestyle changes and suspected COVID-19 symptoms. The factors that demonstrated associations with outcome variables at P values <.20 in the bivariate models were included in multivariate models [63]. To exclude colliders, which may cause multicollinearity, we checked the correlations between them using the Spearman correlation test. The representative factors were selected in the multivariate analysis. The regression coefficient (B), odds ratio (OR), and 95% confidence interval (95% CI) were reported appropriately.

Data were analyzed using IBM SPSS (Version 20.0; IBM Corp). The significance level was set at a P value <.05.

Results

Characteristics of HCWs

Table 2 indicates that, of 5209 HCWs, 905 (17.4%) were aged 41-60 years, 1714 (32.9%) were men, 3915 (75.2%) were ever

married, 2458 (47.2%) could very or fairly easily pay for medication, 4506 (86.5%) had a middle or high social status, 2540 (48.8%) were nurses, 1453 (27.9%) were doctors, 2299 (44.1%) were frontline HCWs, 1999 (38.4%) had epidemic containment experience, 270 (5.2%) had comorbidities, 769 (14.8%) had had suspected COVID-19 symptoms, 5042 (96.8%) ate at an "unchanged or healthier" level, 228 (4.4%) smoked at an "unchanged or more" level, 234 (4.5%) drank at an "unchanged or more" level, and 3553 (68.2%) had an "unchanged or more" physical activity level during the pandemic. The mean (SD) scores of adherence to IPC measures, HL, and eHEALS were 30.6 (6.2), 36.2 (7.3), and 33.1 (4.8), respectively. The HL scores varied by age, gender, ability to pay for medication, social status, type of health care personnel, epidemic containment experience, BMI, suspected COVID-19 symptoms, dietary intake, smoking, and physical activity ($P<.05$). The eHEALS scores varied by gender, ability to pay for medication, type of health care personnel, epidemic containment experience, BMI, suspected COVID-19 symptoms, dietary intake, smoking, and physical activity (Table 2).

Table 2. Characteristics, health literacy, and eHealth literacy among health care workers.

Variables	Total (N=5209)	Health literacy		eHealth Literacy Scale	
	Participants, n (%)	Mean (SD)	*P* value[a]	Mean (SD)	*P* value[a]
Age (years)			.02	N/A[b]	.88
21-40	4304 (82.6)	36.1 (7.2)	N/A	33.1 (4.7)	N/A
41-60	905 (17.4)	36.7 (7.5)	N/A	33.1 (5.0)	N/A
Gender			<.001	N/A	<.001
Women	3495 (67.1)	35.7 (7.0)	N/A	32.8 (4.5)	N/A
Men	1714 (32.9)	37.1 (7.8)	N/A	33.8 (5.3)	N/A
Marital status			.04	N/A	.29
Never married	1294 (24.8)	35.8 (7.0)	N/A	33.0 (4.8)	N/A
Ever married	3915 (75.2)	36.3 (7.4)	N/A	33.2 (4.7)	N/A
Ability to pay for medication			<.001	N/A	<.001
Very or fairly difficult	2751 (52.8)	35.3 (7.4)	N/A	32.8 (4.9)	N/A
Very or fairly easy	2458 (47.2)	37.2 (7.0)	N/A	33.5 (4.6)	N/A
Social status			<.001	N/A	.27
Low	703 (13.5)	35.0 (7.7)	N/A	32.9 (4.7)	N/A
Middle or high	4506 (86.5)	36.3 (7.2)	N/A	33.2 (4.8)	N/A
Type of health care personnel			<.001	N/A	<.001
Other	1216 (23.3)	35.7 (7.6)	N/A	33.1 (5.1)	N/A
Nurse	2540 (48.8)	35.7 (7.2)	N/A	32.7 (4.6)	N/A
Doctor	1453 (27.9)	37.3 (7.1)	N/A	33.9 (4.8)	N/A
Type of health care facility[c]			.77	N/A	.46
Non-frontline	2910 (55.9)	36.2 (7.4)	N/A	33.2 (4.7)	
Frontline	2299 (44.1)	36.1 (7.1)	N/A	33.1 (4.8)	N/A
Epidemic containment experience			<.001	N/A	<.001
No	3210 (61.6)	35.4 (7.2)	N/A	32.9 (4.6)	N/A
Yes	1999 (38.4)	37.4 (7.3)	N/A	33.5 (5.0)	N/A
Comorbidity			.38	N/A	.37
None	4939 (94.8)	36.1 (7.3)	N/A	33.1 (4.8)	N/A
One or more	270 (5.2)	36.5 (7.6)	N/A	32.9 (5.0)	N/A
Suspected COVID-19 symptoms[d]			<.001	N/A	<.001
No	4440 (85.2)	36.4 (7.3)	N/A	33.3 (4.7)	N/A
Yes	769 (14.8)	34.7 (7.2)	N/A	32.3 (5.1)	N/A
Dietary intake[e]			<.001	N/A	.002
Eat less healthy	167 (3.2)	34.2 (6.9)	N/A	32.0 (5.9)	N/A
Unchanged or healthier	5042 (96.8)	36.2 (7.3)	N/A	33.2 (4.7)	N/A
Smoking tobacco[f]			.046	N/A	.03
Never, stopped, or less	4981 (95.6)	36.1 (7.2)	N/A	33.1 (4.7)	N/A
Unchanged or more	228 (4.4)	37.1 (8.6)	N/A	33.8 (5.5)	N/A
Drinking alcohol[f]			.18	N/A	.53
Never, stopped, or less	4975 (95.5)	36.1 (7.2)	N/A	33.1 (4.7)	N/A

Variables	Total (N=5209)	Health literacy		eHealth Literacy Scale	
	Participants, n (%)	Mean (SD)	P value[a]	Mean (SD)	P value[a]
Unchanged or more	234 (4.5)	36.8 (8.3)	N/A	33.3 (5.7)	N/A
Physical activity[f]			<.001	N/A	<.001
Never, stopped, or less	1656 (31.8)	35.2 (7.6)	N/A	32.5 (5.2)	N/A
Unchanged or more	3553 (68.2)	36.6 (7.1)	N/A	33.4 (4.6)	N/A
Adherence to infection prevention and control measures, mean (SD)	30.6 (6.2)	N/A	N/A	N/A	N/A
Health literacy index, mean (SD)	36.2 (7.3)	N/A	N/A	N/A	N/A
eHealth Literacy Scale score, mean (SD)	33.1 (4.8)	N/A	N/A	N/A	N/A

[a]P value: result of one-way analysis of variance (ANOVA) test.

[b]N/A: not applicable.

[c]Frontline areas include the outpatient department, emergency department, isolation areas, imaging and laboratory diagnosis department, and patient administration areas.

[d]Suspected COVID-19 symptoms include common symptoms (fever, cough, dyspnea) and less common symptoms (myalgia, fatigue, sputum production, confusion, headache, sore throat, rhinorrhea, chest pain, hemoptysis, diarrhea, and nausea/vomiting).

[f]Health care workers were asked whether their lifestyle behaviors got worse, better, or were unchanged during the COVID-19 pandemic as compared to before the pandemic.

Psychometric Properties of eHealth Literacy

As shown in Table 3, the KMO value of the overall scale was 0.93 and the values of its 8 items ranged from 0.92 to 0.95. The Bartlett Test of Sphericity value was <0.001, which is a satisfactory level. In addition, the average communality value of 0.76 was satisfactory, demonstrating the accuracy of the approach [59]. Overall, the 8 items of the eHEALS were strongly loaded on one component and explained 76.34% of the scale variance. The factor loading values of 8 items were ranked from 0.78 to 0.92, as shown in Table 3. The correlations between each item and the scale range from 0.80 to 0.84, indicating satisfactory convergent validity [50,51]. In addition, the correlation between eHEALS and HL scores was moderate (ρ=0.42), providing evidence of criterion validity [60]. Furthermore, the Cronbachα value of .95 indicated a high level of internal consistency. There was no significant floor effect, with 0.70% of participants having the lowest potential response. A marginal ceiling effect was found, with 16.1% of participants having the highest potential response (slightly higher than the recommended percentage of <15%; Table 3).

Table 3. Construct, convergent, and criterion validity, internal consistency, and floor and ceiling effects of the 8-item eHealth Literacy Scale (N=5209).

Construct validity, factor loadings	Values
eHealth Literacy scale items	
I know what health resources are available on the internet	0.85
I know where to find helpful health resources on the internet	0.90
I know how to find helpful health resources on the internet	0.92
I know how to use the internet to answer my questions about health	0.90
I know how to use the health information I find on the internet to help me	0.92
I have the skills I need to evaluate the health resources I find on the internet	0.89
I have the skills needed to tell high-quality health resources from low-quality health resources on the internet	0.82
I feel confident in using information from the internet to make health decisions	0.78
Percentage of variance, %	76.34
Item-scale convergent validity, mean of ρ^a (range)	0.83 (0.80-0.84)
Criterion validity, correlation with health literacy, ρ^b	0.42
Internal consistency, Cronbach α	.95
Floor effect, %	0.70
Ceiling effect, %	16.10

[a]ρ: Spearman correlation coefficient.

[b]ρ: Pearson correlation coefficient.

Determinants of HL and eHealth Literacy

The week correlations among independent variables (ρ <0.30) suggest that there is no collider which might affect the results (Table 1 in Multimedia Appendix 1). Results of the multivariate analysis shown in Table 4 indicated that HCWs with higher HL scores were men (unstandardized regression coefficient [B]=1.01, 95% CI 0.57-1.45, *P*<.001), those with a very or fairly easy ability to pay for medication (B=1.65, 95% CI 1.25-2.05, *P*<.001), those with a middle or high social status (B=0.586, 95% CI 0.003-1.169, *P*=.049), doctors (B=1.29, 95% CI 0.73-1.84, *P*<.001), and those with epidemic containment experience (B=1.96, 95% CI 1.56-2.37, *P*<.001), as compared to their counterparts, respectively.

HCWs with higher eHEALS scores were men (B=0.72, 95% CI 0.43-1.00, *P*<.001), those with a very or fairly easy ability to pay for medication (B=0.60, 95% CI 0.34-0.86, *P*<.001), doctors (B=0.56, 95% CI 0.20-0.93, *P*=.003), and those with epidemic containment experience (B=0.64, 95% CI 0.38-0.91, *P*<.001), as compared to their counterparts, respectively (Table 4).

Table 4. Determinants of health literacy and eHealth literacy among health care workers (N=5209).

Variables	Health literacy				eHealth literacy			
	Bivariate		Multivariate		Bivariate		Multivariate	
	B[a] (95% CI)	P value	B (95% CI)	P value	B (95% CI)	P value	B (95% CI)	P value
Age (years)								
21-40	Reference	N/A[b]	Reference	N/A	Reference	N/A	Reference	N/A
41-60	0.62 (0.10 to 1.14)	.02	–0.05 (–0.58 to 0.49)	.86	–0.03 (–0.37 to 0.31)	.88	–0.30 (–0.65 to 0.04)	.08
Gender								
Female	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Male	1.36 (0.94 to 1.78)	<.001	1.01 (0.57 to 1.45)	<.001	0.99 (0.72 to 1.27)	<.001	0.72 (0.43 to 1.00)	<.001
Marital status								
Never married	Reference	N/A	Reference	N/A	Reference	N/A	N/A	N/A
Ever married	0.47 (0.01 to 0.93)	.04	0.34 (–0.12 to 0.81)	.15	0.16 (–0.14 to 0.46)	.29	N/A	N/A
Ability to pay for medication								
Very or fairly difficult	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Very or fairly easy	1.9 (1.51 to 2.29)	<.001	1.65 (1.25 to 2.05)	<.001	0.72 (0.46 to 0.98)	<.001	0.60 (0.34 to 0.86)	<.001
Social status								
Low	Reference	N/A	Reference	N/A	Reference	N/A	N/A	N/A
Middle or high	1.29 (0.72 to 1.87)	<.001	0.586 (0.003 to 1.169)	.049	0.22 (–0.16 to 0.59)	.27	N/A	N/A
Type of health care personnel								
Other	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Nurse	0.09 (–0.41 to 0.58)	.73	0.18 (–0.32 to 0.68)	.48	–0.40 (–0.72 to –0.07)	.02	–0.32 (–0.65 to 0.01)	.06
Doctor	1.68 (1.13 to 2.24)	<.001	1.29 (0.73 to 1.84)	<.001	0.76 (0.39 to 1.12)	<.001	0.56 (0.20 to 0.93)	.003
Type of health care facility[c]								
Non-frontline	Reference	N/A	N/A	N/A	Reference	N/A	N/A	N/A
Frontline	–0.06 (–0.46 to 0.34)	.77	N/A	N/A	–0.10 (–0.36 to 0.16)	.46	N/A	N/A
Epidemic containment experience								
No	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Yes	1.95 (1.54 to 2.35)	<.001	1.96 (1.56 to 2.37)	<.001	0.55 (0.29 to 0.82)	<.001	0.64 (0.38 to 0.91)	<.001
Comorbidity								
None	Reference	N/A	Reference	N/A	Reference	N/A	N/A	N/A
One or more	0.40 (–0.49 to 1.29)	.378	N/A	N/A	–0.27 (–0.85 to 0.32)	.369	N/A	N/A

[a]B: unstandardized regression coefficient.

[b]N/A: not applicable.

[c]Frontline areas are the outpatient department, emergency department, isolation areas, imaging and laboratory diagnosis department, and patient administration areas.

XSL•FO
RenderX

Association Between HL and eHealth Literacy and Adherence to IPC Procedures

According to the results of a multivariate linear regression analysis (Table 5), higher HL (B=0.13, 95% CI 0.10-0.15, *P*<.001) and eHEALS scores (B=0.22, 95% CI 0.19-0.26,

P<.001) were found to be associated with better adherence to IPC measures after adjusting for age, gender, ability to pay for medication, social status, profession, type of health care facility, epidemic containment experience, and comorbidities. These adjusted factors showed the associations with adherence to IPC measures at *P*<.20 (Table 2 in Multimedia Appendix 1).

Table 5. Associations of health literacy and eHealth literacy with adherence to infection prevention and control measures, lifestyle changes, and suspected COVID-19 symptoms among health care workers (N=5209).

Variables	Adherence to IPC measures[a]		Dietary intake[b]		Smoking tobacco[c]		Drinking alcohol[d]		Physical activity[e]		Suspected COVID-19 symptoms[f]	
	B[g] (95% CI)	*P* value	OR[h] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Health literacy, 1-score increment												
Bivariate model	0.13 (0.11-0.15)	<.001	1.04 (1.02-1.06)	<.001	1.02 (1.00-1.04)	.046	1.01 (0.99-1.03)	.18	1.03 (1.02-1.04)	<.001	0.97 (0.96-0.98)	<.001
Multivariate model	0.13 (0.10-0.15)	<.001	1.04 (1.01-1.06)	.001	1.01 (0.99-1.03)	.34	1.01 (0.99-1.02)	.47	1.03 (1.02-1.03)	<.001	0.97 (0.96-0.98)	<.001
eHealth Literacy Scale, 1-score increment												
Bivariate model	0.22 (0.18-0.25)	<.001	1.04 (1.02-1.07)	.002	1.03 (1-1.07)	.03	1.01 (0.98-1.04)	.53	1.04 (1.03-1.05)	<.001	0.96 (0.95-0.98)	<.001
Multivariate model	0.22 (0.19-0.26)	<.001	1.04 (1.02-1.07)	.002	1.01 (0.99-1.04)	.30	0.99 (0.97-1.02)	.69	1.04 (1.03-1.05)	<.001	0.96 (0.95-0.98)	<.001

[a]Adherence to infection prevention and control procedures; adjusted for age, gender, ability to pay for medication, social status, type of health care personnel, type of health care facility, epidemic containment experience, and comorbidity in the multivariate model.

[b]Adjusted for age, gender, marital status, ability to pay for medication, and social status in the multivariate model. The reference group is "less healthy diet" and the test group is "unchanged or healthier diet."

[c]Adjusted for age, gender, marital status, social status, and type of health care personnel in the multivariate model. The reference group is "never, stopped, or less smoking," and the test group is "unchanged or more smoking."

[d]Adjusted for age, gender, social status, type of health care personnel, and epidemic containment experience in the multivariate model. The reference group is "never, stopped, or less drinking," and the test group is "unchanged or more drinking."

[e]Adjusted for age, gender, ability to pay for medication, social status, type of health care personnel, type of health care facility, and epidemic containment experience in the multivariate model. The reference group is "never, stopped, or less physical activity," and the test group is "unchanged or more physical activity."

[f]Suspected COVID-19 symptoms include common symptoms (fever, cough, dyspnea) and less common symptoms (myalgia, fatigue, sputum production, confusion, headache, sore throat, rhinorrhea, chest pain, hemoptysis, diarrhea, and nausea/vomiting). Adjusted for age, gender, marital status, ability to pay for medications, social status, type of health care personnel, and comorbidity in the multivariate model.

[g]B: unstandardized regression coefficient.

[h]OR: odds ratio.

Association Between HL and eHealth Literacy and Lifestyle Changes

The results of the multivariate logistic regression analysis shown in Table 5 indicated that HCWs with higher HL (OR 1.04, 95% CI 1.01-1.06, *P*=.001) and higher eHEALS scores (OR 1.04, 95% CI 1.02-1.07, *P*=.002) had a higher likelihood of eating at an "unchanged or healthier" level after being adjusted for age, gender, marital status, ability to pay for medication, and social status. These adjusted factors had associations with eating behavior changes at *P*<.20 (Table 2 in Multimedia Appendix

1). In addition, HCWs with higher HL scores (OR 1.03, 95% CI 1.02-1.03, *P*<.001) and higher eHEALS scores (OR 1.04, 95% CI 1.03-1.05, *P*<.001) had a higher likelihood of doing physical activity at an "unchanged or more" level after adjusting for age, gender, ability to pay for medication, social status, profession, type of health care facility, and epidemic containment experience. These adjusted factors had associations with physical activity at *P*<.20 (Table 2 in Multimedia Appendix 1). No association was found between HL and eHEALS scores and smoking or drinking behaviors.

Associations of HL and eHealth Literacy With Suspected COVID-19 Symptoms

As shown in Table 5, HCWs with higher HL scores (OR 0.97, 95% CI 0.96-0.98, *P*<.001) and higher eHEALS scores (OR 0.96, 95% CI 0.95-0.98, *P*<.001) had a lower likelihood of having suspected COVID-19 symptoms after adjusting for age, gender, marital status, ability to pay for medication, social status, profession, and comorbidities. These adjusted factors had associations with suspected COVID-19 symptoms at *P*<.20 (Table 2 in Multimedia Appendix 1).

Discussion

Principal Findings

The eHEALS questionnaire was found to be valid and reliable for assessing the eHealth literacy of HCWs. The tool was found to have satisfactory construct validity, convergent validity, criterion validity, and reliability. There was no flooring effect. A marginal ceiling effect was found. A high percentage of participants with the highest possible score was also observed in a previous study that validated the digital health literacy scale [64]. This could be because HCWs have been recognized as a group that has a higher standard of health knowledge and skills than the average person. Therefore, it is reasonable that many participants had a higher or maximum score.

In this study, male HCWs had higher HL and eHEALS scores. This finding is inconsistent with a previous study conducted among the general population in Europe [65], in which men had lower HL scores than women. Another study conducted among the general population in Taiwan showed that men also had lower scores of HL and digital health diet literacy than women [66]. In addition, eHealth literacy was found to be higher in women than men in the general population in Hungary [67] and in health care professionals in Ethiopia [68]. This inconsistency could be explained by the fact that although the total HCW population had fewer males (32.9% versus 67.1% female), the proportion of male doctors (797/1453 or 54.9%) was higher than that of female doctors (656/1453 or 45.1%).

Doctors had higher HL and eHEALS scores than other HCWs in this study. Similarly, HCWs with epidemic containment experience had higher HL and eHEALS scores in this study. Among HCWs, doctors receive longer professional training and they have been recognized as the group with the highest level of ability to find, understand, justify, and use health-related information. In addition, they are involved in educating and counseling other HCWs and patients [69]. HCWs who were involved in containing previous outbreaks or epidemics logically had better health knowledge and skills that made their HL and eHEALS scores higher. HL was found to be a strategic approach for improving communication between patients and HCWs [41], further contributing to the containment of the pandemic and its consequences [45].

HCWs with a better ability to pay for medication had higher HL and eHEALS scores. A positive association between the ability to pay for medication was prominently found in many previous studies [51,52,66,70,71]. In addition, HCWs with middle or high social status had higher HL scores. This was

similar to the findings of previous studies conducted in Asia [51,52,66,70-72] and Europe [65]. This implies that socioeconomic factors are independent indictors that should be taken into consideration in intervention development to improve HL and eHEALS scores.

In this study, HCWs with higher HL or eHEALS scores had better adherence to IPC procedures. This is the first study investigating these associations. In the literature, higher HL scores were found to be associated with better adherence to therapies among people with chronic diseases [73-76]. Receiving COVID-19–related information online during the pandemic could be beneficial for various preventive behaviors [77]. Therefore, more efforts are needed to increase the distribution of accurate information and encourage appropriate behaviors [78]. HCWs should be aware of and strictly adhere to the guidelines for IPC to contain the COVID-19 pandemic.

Our study found that HCWs with higher HL and eHEALS scores had a lower likelihood of suspected COVID-19 symptoms. It has been noted that individuals with higher HL scores had better health status [71,79] and well-being [80,81]. This could be further explained by the fact that higher HL and eHEALS scores were associated with healthier eating and physical activity behaviors in this study. In a previous study, we also found that medical and nursing students with higher digital healthy diet literacy had healthier eating behaviors during the pandemic [82]. In addition, eHealth literacy was found to be associated with key issues related to seeking out exercise regime information and maintaining a healthy diet [83]. In addition, it has been reported that people with higher HL scores had healthier behaviors (eg, exercise, balanced diet) [66,84], further protecting and improving their health and well-being [85].

This study has some limitations. First, the study was conducted online and the suspected COVID-19 symptoms were self-reported; therefore, we cannot confirm COVID-19 cases and exclude them from our study. Fortunately, there were no new confirmed cases during the data collection period [86]. Second, during the lockdown period, people were ordered to stay at home and only go out to buy food or medicine, for emergencies, or to work at businesses producing and providing essential goods and services. This may cause a measurement bias in the application domain of HL, as the questionnaire cannot distinguish between an individual's ability and restrictions during lockdown (eg, on questions such as "on a scale from very easy to very difficult, how easy would you say it is to join a sports club or exercise class if you want to?"). Third, a cross-sectional design was used, which cannot provide evidence of the causal relationships among studied variables. Furthermore, the convenient sampling technique used may limit the generalizability of the findings. In addition, the differences between some categories were not examined due to small sample sizes (eg, widowed/separated/divorced participants [140/5209, 2.7%] or those with high social status 63/5209, 1.2%]). However, we surveyed a relatively large sample of HCWs from 15 hospitals and health centers across Vietnam, which can help with exploring associations in future research and provides evidence for potential interventions.

Conclusions

The eHEALS questionnaire is a valid and reliable tool for assessing eHealth literacy among HCWs. HL and eHealth literacy were significantly higher in men, those with better ability to pay for medication, doctors, and those with previous epidemic containment experience. Both HL and eHealth literacy were associated with better adherence to IPC procedures, healthier lifestyles (eg, healthier eating behavior and more physical activity during the pandemic), and a lower likelihood of having suspected COVID-19 symptoms. Integrative and multidisciplinary approaches are required to improve HCWs' HL and eHealth literacy, which could help improve adherence to IPC measures, promote healthy behaviors, and protect the health of HCWs. This would further contribute to containing the COVID-19 pandemic and minimizing its consequences.

Authors' Contributions

BND, TVT, and TVD analyzed the data and drafted the manuscript. BND, TVT, DTP, HCN, TTPN, HCN, THH, HKD, MVT, TVD, HQN, TTN, NPTN, CQT, KVT, TTD, HXP, LVN, KTN, PWC, and TVD contributed to conceptualization, investigation, methodology, validation, and manuscript revision. BND, TVT, DTP, HCN, TTPN, HCN, THH, HKD, MVT, TVD, HQN, TTN, NPTN, CQT, KVT, TTD, HXP, LVN, KTN, and TVD conducted data curation. All authors gave final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 38 KB - jmir_v22i11e22894_app1.docx]

References

1. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]
2. Bassetti M, Vena A, Giacobbe DR. The novel Chinese coronavirus (2019-nCoV) infections: Challenges for fighting the storm. Eur J Clin Invest 2020 Mar;50(3):e13209. [doi: 10.1111/eci.13209] [Medline: 32003000]
3. Phelan AL, Katz R, Gostin LO. The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance. JAMA 2020 Feb 25;323(8):709-710 [FREE Full text] [doi: 10.1001/jama.2020.1097] [Medline: 31999307]
4. Rubin R. The Challenge of Preventing COVID-19 Spread in Correctional Facilities. JAMA 2020 May 12;323(18):1760-1761. [doi: 10.1001/jama.2020.5427] [Medline: 32259189]
5. Xiang Y, Jin Y, Cheung T. Joint International Collaboration to Combat Mental Health Challenges During the Coronavirus Disease 2019 Pandemic. JAMA Psychiatry 2020 Oct 01;77(10):989-990. [doi: 10.1001/jamapsychiatry.2020.1057] [Medline: 32275289]
6. Campion J, Javed A, Sartorius N, Marmot M. Addressing the public mental health challenge of COVID-19. The Lancet Psychiatry 2020 Aug;7(8):657-659. [doi: 10.1016/s2215-0366(20)30240-6]
7. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
8. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health 2020 Aug;8(8):e1003-e1017 [FREE Full text] [doi: 10.1016/S2214-109X(20)30264-3] [Medline: 32553130]
9. Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. Lancet 2020 May 30;395(10238):1715-1725 [FREE Full text] [doi: 10.1016/S0140-6736(20)30854-0] [Medline: 32405103]
10. Moradian N, Ochs HD, Sedikies C, Hamblin MR, Camargo CA, Martinez JA, et al. The urgent need for integrated science to fight COVID-19 pandemic and beyond. J Transl Med 2020 May 19;18(1):205 [FREE Full text] [doi: 10.1186/s12967-020-02364-2] [Medline: 32430070]
11. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]

XSL•FO
RenderX

12. Nazir M, Hussain I, Tian J, Akram S, Mangenda Tshiaba S, Mushtaq S, et al. A Multidimensional Model of Public Health Approaches Against COVID-19. Int J Environ Res Public Health 2020 May 26;17(11):3780 [FREE Full text] [doi: 10.3390/ijerph17113780] [Medline: 32466581]

13. Betsch C. How behavioural science data helps mitigate the COVID-19 crisis. Nat Hum Behav 2020 May 27;4(5):438-438 [FREE Full text] [doi: 10.1038/s41562-020-0866-1] [Medline: 32221514]

14. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May 30;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

15. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250 [FREE Full text] [doi: 10.1016/S2468-2667(20)30086-4] [Medline: 32302535]

16. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]

17. PM orders strict nationwide social distancing rules, starting April 1. Vietnam Law Magazine. 2020. URL: https://vietnamlawmagazine.vn/pm-orders-strict-nationwide-social-distancing-rules-starting-april-1-27108.html [accessed 2020-03-31]

18. Hao F, Tan W, Jiang L, Zhang L, Zhao X, Zou Y, et al. Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. Brain Behav Immun 2020 Jul;87:100-106 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.069] [Medline: 32353518]

19. Ozamiz-Etxebarria N, Idoiaga Mondragon N, Dosil Santamaría M, Picaza Gorrotxategi M. Psychological Symptoms During the Two Stages of Lockdown in Response to the COVID-19 Outbreak: An Investigation in a Sample of Citizens in Northern Spain. Front Psychol 2020;11:1491 [FREE Full text] [doi: 10.3389/fpsyg.2020.01491] [Medline: 32625157]

20. Di Renzo L, Gualtieri P, Pivari F, Soldati L, Attinà A, Cinelli G, et al. Eating habits and lifestyle changes during COVID-19 lockdown: an Italian survey. J Transl Med 2020 Jun 08;18(1):229 [FREE Full text] [doi: 10.1186/s12967-020-02399-5] [Medline: 32513197]

21. Ammar A, Brach M, Trabelsi K, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of COVID-19 Home Confinement on Eating Behaviour and Physical Activity: Results of the ECLB-COVID19 International Online Survey. Nutrients 2020 May 28;12(6):1583 [FREE Full text] [doi: 10.3390/nu12061583] [Medline: 32481594]

22. Sidor A, Rzymski P. Dietary Choices and Habits during COVID-19 Lockdown: Experience from Poland. Nutrients 2020 Jun 03;12(6):1657 [FREE Full text] [doi: 10.3390/nu12061657] [Medline: 32503173]

23. Rodríguez-Pérez C, Molina-Montes E, Verardo V, Artacho R, García-Villanova B, Guerra-Hernández EJ, et al. Changes in Dietary Behaviours during the COVID-19 Outbreak Confinement in the Spanish COVIDiet Study. Nutrients 2020 Jun 10;12(6):1730 [FREE Full text] [doi: 10.3390/nu12061730] [Medline: 32531892]

24. Wardell J, Kempe T, Rapinda K, Single A, Bilevicius E, Frohlich J, et al. Drinking to Cope During COVID-19 Pandemic: The Role of External and Internal Factors in Coping Motive Pathways to Alcohol Use, Solitary Drinking, and Alcohol Problems. Alcohol Clin Exp Res 2020 Sep 01:1. [doi: 10.1111/acer.14425] [Medline: 32870516]

25. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]

26. Sesagiri Raamkumar A, Tan S, Wee H. Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]

27. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]

28. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

29. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]

30. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]

31. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]

32. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]

33. Shimizu K. 2019-nCoV, fake news, and racism. Lancet 2020 Feb 29;395(10225):685-686 [FREE Full text] [doi: 10.1016/S0140-6736(20)30357-3] [Medline: 32059801]

34. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]

35. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]

36. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

37. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897. [doi: 10.2196/18897] [Medline: 32325426]

38. Ko N, Lu W, Chen Y, Li D, Wang P, Hsu S, et al. COVID-19-related information sources and psychological well-being: An online survey study in Taiwan. Brain Behav Immun 2020 Jul;87:153-154 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.019] [Medline: 32389702]

39. Mackert M, Ball J, Lopez N. Health literacy awareness training for healthcare workers: improving knowledge and intentions to use clear communication techniques. Patient Educ Couns 2011 Dec;85(3):e225-e228. [doi: 10.1016/j.pec.2011.02.022] [Medline: 21474264]

40. Yunyongying P, Rich M, Jokela J. Patient-Centered Performance Metrics. JAMA 2019 May 14;321(18):1829. [doi: 10.1001/jama.2019.2056] [Medline: 31087017]

41. McKinn S, Linh DT, Foster K, McCaffery K. Communication Between Health Workers and Ethnic Minorities in Vietnam. Health Lit Res Pract 2017 Oct 01;1(4):e163-e172 [FREE Full text] [doi: 10.3928/24748307-20170629-01] [Medline: 31294262]

42. Tavakoly Sany SB, Behzhad F, Ferns G, Peyman N. Communication skills training for physicians improves health literacy and medical outcomes among patients with hypertension: a randomized controlled trial. BMC Health Serv Res 2020 Jan 23;20(1):60 [FREE Full text] [doi: 10.1186/s12913-020-4901-8] [Medline: 31973765]

43. Wang M, Hung L, Lo Y. Glycemic control in type 2 diabetes: role of health literacy and shared decision-making. Patient Prefer Adherence 2019;13:871-879 [FREE Full text] [doi: 10.2147/PPA.S202110] [Medline: 31213781]

44. Redelmeier DA, Shafir E. Pitfalls of judgment during the COVID-19 pandemic. Lancet Public Health 2020 Jun;5(6):e306-e308. [doi: 10.1016/S2468-2667(20)30096-7] [Medline: 32334647]

45. Sentell T, Vamos S, Okan O. Interdisciplinary Perspectives on Health Literacy Research Around the World: More Important Than Ever in a Time of COVID-19. Int J Environ Res Public Health 2020 Apr 26;17(9):3010 [FREE Full text] [doi: 10.3390/ijerph17093010] [Medline: 32357457]

46. Govt extends social distancing for at least one week in 28 localities. VGP News. 2020. URL: http://news.chinhphu.vn/Home/Govt-extends-social-distancing-for-at-least-one-week-in-28-localities/20204/39735.vgp [accessed 2020-04-15]

47. Govt to gradually ease COVID-19 control measure in cautious manner. VGP News. 2020. URL: http://primeminister.chinhphu.vn/Home/Govt-to-gradually-ease-COVID19-control-measure-in-cautious-manner/20204/3759.vgp [accessed 2020-04-20]

48. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol 2011 Mar 15;173(6):676-682. [doi: 10.1093/aje/kwq433] [Medline: 21330339]

49. Charlson ME, Pompei P, Ales KL, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]

50. Duong TV, Aringazina A, Kayupova G, Nurjanah, Pham TV, Pham KM, et al. Development and Validation of a New Short-Form Health Literacy Instrument (HLS-SF12) for the General Public in Six Asian Countries. Health Lit Res Pract 2019 Apr;3(2):e91-e102 [FREE Full text] [doi: 10.3928/24748307-20190225-01] [Medline: 31294310]

51. Duong TV, Nguyen TTP, Pham KM, Nguyen KT, Giap MH, Tran TDX, et al. Validation of the Short-Form Health Literacy Questionnaire (HLS-SF12) and Its Determinants among People Living in Rural Areas in Vietnam. Int J Environ Res Public Health 2019 Sep 11;16(18):3346 [FREE Full text] [doi: 10.3390/ijerph16183346] [Medline: 31514271]

52. Van Hoa H, Giang HT, Vu PT, Van Tuyen D, Khue PM. Factors Associated with Health Literacy among the Elderly People in Vietnam. Biomed Res Int 2020 Mar 27;2020:3490635-3490637. [doi: 10.1155/2020/3490635] [Medline: 32309429]

53. Nguyen HC, Nguyen MH, Do BN, Tran CQ, Nguyen TTP, Pham KM, et al. People with Suspected COVID-19 Symptoms Were More Likely Depressed and Had Lower Health-Related Quality of Life: The Potential Benefit of Health Literacy. J Clin Med 2020 Mar 31;9(4):965 [FREE Full text] [doi: 10.3390/jcm9040965] [Medline: 32244415]

54. Nguyen HT, Do BN, Pham KM, Kim GB, Dam HT, Nguyen TT, et al. Fear of COVID-19 Scale-Associations of Its Scores with Health Literacy and Health-Related Behaviors among Medical Students. Int J Environ Res Public Health 2020 Jun 11;17(11):4164. [doi: 10.3390/ijerph17114164] [Medline: 32545240]

55. HLS-EU Consortium. Comparative report of health literacy in eight EU member states. The European Health Literacy Project. 2009. URL: https://www.healthliteracyeurope.net/hls-eu [accessed 2012-10-22]

56. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

57. World Health Organization. Risk assessment and management of exposure of health care workers in the context of COVID-19: Interim guidance. 2020. URL: https://www.who.int/publications/i/item/risk-assessment-and-management-of-exposure-of-health-care-workers-in-the-context-of-covid-19-interim-guidance [accessed 2020-03-19]

58. Editorial Team. Overview of novel coronavirus (2019-nCoV). BMJ Best Practice. 2020. URL: https://bestpractice.bmj.com/topics/en-gb/3000165 [accessed 2020-02-10]

59. Field A. Discovering statistics using IBM SPSS statistics. Washington DC, USA: Sage Publications Ltd; 2013.

60. Kimberlin C, Winterstein A. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm 2008 Dec 01;65(23):2276-2284. [doi: 10.2146/ajhp070364] [Medline: 19020196]

61. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]

62. Cronbach LJ, Shavelson RJ. My Current Thoughts on Coefficient Alpha and Successor Procedures. Educational and Psychological Measurement 2016 Jul 02;64(3):391-418. [doi: 10.1177/0013164404266386]

63. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. Am J Epidemiol 1993 Dec 01;138(11):923-936. [doi: 10.1093/oxfordjournals.aje.a116813] [Medline: 8256780]

64. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

65. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec 05;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]

66. Van Duong T, Chiu C, Lin C, Chen YC, Wong TC, Chang P, et al. E-healthy diet literacy scale and its relationship with behaviors and health outcomes in Taiwan. Health Promot Int 2020 Apr 08:E. [doi: 10.1093/heapro/daaa033] [Medline: 32267935]

67. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jun 16;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]

68. Shiferaw KB, Mehari EA. Internet use and eHealth literacy among health-care professionals in a resource limited setting: a cross-sectional survey. Adv Med Educ Pract 2019;10:563-570 [FREE Full text] [doi: 10.2147/AMEP.S205414] [Medline: 31440113]

69. Palesy D, Jakimowicz S. Health literacy training for Australian home care workers: Enablers and barriers. Home Health Care Serv Q 2019 Apr 20;38(2):80-95. [doi: 10.1080/01621424.2019.1604458] [Medline: 31007141]

70. Duong T, Sørensen K, Pelikan JM, Van den Broucke S, Lin I, Lin Y, et al. Health-related behaviors moderate the association between age and self-reported health literacy among Taiwanese women. Women Health 2018 Jul 06;58(6):632-646. [doi: 10.1080/03630242.2017.1333074] [Medline: 28537772]

71. Kim D, Chang H. Key functional characteristics in designing and operating health information websites for user satisfaction: an application of the extended technology acceptance model. Int J Med Inform 2007;76(11-12):790-800. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]

72. Duong TV, Aringazina A, Baisunova G, Nurjanah, Pham TV, Pham KM, et al. Measuring health literacy in Asia: Validation of the HLS-EU-Q47 survey tool in six Asian countries. J Epidemiol 2017 Feb;27(2):80-86 [FREE Full text] [doi: 10.1016/j.je.2016.09.005] [Medline: 28142016]

73. Elisabeth Stømer U, Klopstad Wahl A, Gunnar Gøransson L, Hjorthaug Urstad K. Health Literacy in Kidney Disease: Associations with Quality of Life and Adherence. J Ren Care 2020 Jun 16;46(2):85-94. [doi: 10.1111/jorc.12314] [Medline: 31950601]

74. Huang Y, Shiyanbola OO, Chan H. A path model linking health literacy, medication self-efficacy, medication adherence, and glycemic control. Patient Educ Couns 2018 Nov;101(11):1906-1913. [doi: 10.1016/j.pec.2018.06.010] [Medline: 29935739]

75. Lor M, Koleck TA, Bakken S, Yoon S, Dunn Navarra A. Association Between Health Literacy and Medication Adherence Among Hispanics with Hypertension. J Racial Ethn Health Disparities 2019 Jun 3;6(3):517-524 [FREE Full text] [doi: 10.1007/s40615-018-00550-z] [Medline: 30607576]

76. Skoumalova I, Kolarcik P, Madarasova Geckova A, Rosenberger J, Majernikova M, Klein D, et al. Is Health Literacy of Dialyzed Patients Related to Their Adherence to Dietary and Fluid Intake Recommendations? Int J Environ Res Public Health 2019 Nov 05;16(21):4295 [FREE Full text] [doi: 10.3390/ijerph16214295] [Medline: 31694265]

77. Li S, Feng B, Liao W, Pan W. Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]

78. Alsan M, Stantcheva S, Yang D, Cutler D. Disparities in Coronavirus 2019 Reported Incidence, Knowledge, and Behavior Among US Adults. JAMA Netw Open 2020 Jun 01;3(6):e2012403 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12403] [Medline: 32556260]

79. Duong VT, Lin I, Sorensen K, Pelikan JM, Van Den Broucke S, Lin Y, et al. Health Literacy in Taiwan: A Population-Based Study. Asia Pac J Public Health 2015 Nov 29;27(8):871-880. [doi: 10.1177/1010539515607962] [Medline: 26419635]

80. Watson R. Europeans with poor "health literacy" are heavy users of health services. BMJ 2011 Nov 29;343(nov29 4):d7741-d7741. [doi: 10.1136/bmj.d7741] [Medline: 22127773]

81. Greenhalgh T. Health literacy: towards system level solutions. BMJ 2015 Feb 24;350(feb24 13):h1026-h1026. [doi: 10.1136/bmj.h1026] [Medline: 25712067]

82. Duong TV, Pham KM, Do BN, Kim GB, Dam HTB, Le VT, et al. Digital Healthy Diet Literacy and Self-Perceived Eating Behavior Change during COVID-19 Pandemic among Undergraduate Nursing and Medical Students: A Rapid Online Survey. Int J Environ Res Public Health 2020 Sep 30;17(19):7185 [FREE Full text] [doi: 10.3390/ijerph17197185] [Medline: 33008102]

83. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth Literacy and Health Behaviors Affecting Modern College Students: A Pilot Study of Issues Identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]

84. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]

85. Owen L, Corfe B. The role of diet and nutrition on mental health and wellbeing. Proc Nutr Soc 2017 Nov;76(4):425-426. [doi: 10.1017/S0029665117001057] [Medline: 28707609]

86. Ministry of Health. Coronavirus disease (COVID-19) outbreak in Vietnam. 2020. URL: https://ncov.moh.gov.vn/ [accessed 2020-05-04]

Abbreviations

B: unstandardized regression coefficient
eHEALS: eHealth Literacy Scale
HCWs: health care workers
HL: health literacy
HLS-SF12: 12-item short-form health literacy questionnaire
IPC: infection prevention and control
KMO: Kaiser-Meyer Olkin Measure
PCA: principal component analysis
PPE: personal protective equipment
WHO: World Health Organization

Original Paper

Relationship Between COVID-19 Infection and Risk Perception, Knowledge, Attitude, and Four Nonpharmaceutical Interventions During the Late Period of the COVID-19 Epidemic in China: Online Cross-Sectional Survey of 8158 Adults

Hong Xu[1,2], PhD; Yong Gan[1], PhD; Daikun Zheng[3], MBBS; Bo Wu[4], MBBS; Xian Zhu[2], MBBS; Chang Xu[2], BBA; Chenglu Liu[2], BBA; Zhou Tao[2], BBA; Yaoyue Hu[2], PhD; Min Chen[2], MM; Mingjing Li[2], MSc; Zuxun Lu[1], PhD; Jack Chen[2,5], PhD

[1]School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhai, China

[2]School of Public Health and Management, Chongqing Medical University, Chongqing, China

[3]Chongqing Three Gorges Medical College, Chongqing, China

[4]Wanzhou District Center for Disease Control and Prevention, Chongqing, China

[5]Ingham Institute & University of New South Wales, Sydney, Australia

Corresponding Author:
Zuxun Lu, PhD
School of Public Health
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Hubei, PRC
Wuhai, 430030
China
Phone: 86 27 8754 2101
Email: zuxunlu@yahoo.com

Abstract

Background: So far, there have been no published population studies on the relationship between a COVID-19 infection and public risk perception, information source, knowledge, attitude, and behaviors during the COVID-19 outbreak in China.

Objective: This study aims to understand the relationships between COVID-19 infection; four personal nonpharmaceutical interventions (NPIs; handwashing, proper coughing habits, social distancing, and mask wearing); and public risk perception, knowledge, attitude, and other social demographic variables.

Methods: An online survey of 8158 Chinese adults between February 22 and March 5, 2020, was conducted. Bivariate associations between categorical variables were examined using Fisher exact test. We also explored the determinants of four NPIs as well as their association with COVID-19 infection using logistic regression.

Results: Of 8158 adults included, 57 (0.73%) were infected with COVID-19. The overwhelming majority of respondents showed a positive attitude (n=8094, 99.2%), positive risk perception (n=8146, 99.9%), and high knowledge levels that were among the strongest predictors of the four adopted NPIs (handwashing: n=7895, 96.8%; proper coughing: 5997/6444, 93.1%; social distancing: n=7104/8158, 87.1%; and mask wearing: 5011/5120, 97.9%). There was an increased risk of COVID-19 infection for those who did not wash their hands (2.28% vs 0.65%; risk ratio [RR] 3.53, 95% CI 1.53-8.15; *P*=.009), did not practice proper coughing (1.79% vs 0.73%; RR 2.44, 95% CI 1.15-5.15; *P*=.03), did not practice social distancing (1.52% vs 0.58%; RR 2.63, 95% CI 1.48-4.67; *P*=.002), and did not wear a mask (7.41% vs 0.6%; RR 12.38, 95% CI 5.81-26.36; *P*<.001). For those who did practice all other three NPIs, wearing a mask was associated with a significantly reduced risk of infection compared to those who did not wear a mask (0.6% vs 16.7%; *P*=.04). Similarly, for those who did not practice all or part of the other three NPIs, wearing a mask was also associated with a significantly reduced risk of infection. In a penalized logistic regression model including all four NPIs, wearing a mask was the only significant predictor of COVID-19 infection among the four NPIs (odds ratio 7.20, 95% CI 2.24-23.11; *P*<.001).

Conclusions: We found high levels of risk perception, positive attitude, desirable knowledge, as well as a high level of adopting the four NPIs. The relevant knowledge, risk perception, and attitude were strong predictors of adapting the four NPIs. Mask wearing, among the four personal NPIs, was the most effective protective measure against COVID-19 infection, with added preventive effect among those who practiced all or part of the other three NPIs.

Introduction

The unprecedented COVID-19 global pandemic [1] has changed the way our society operates. The total confirmed cases and deaths worldwide increased at an alarming rate [2]. The availability of an effective vaccine may still be many months away [3,4], and there is no consensus on the use of antiviral drugs and other therapeutic agents [5,6]. Meanwhile, the best hope for reducing mortality is societal preventative measures and providing timely and optimal critical care. As the list of countries in the grip of the rapid spread of COVID-19 is growing, many countries are, or will be, at the brink of further overwhelmed health care systems. Many countries have strengthened their nonpharmaceutical interventions (NPIs) to flatten the curve and reduce casualties [7]. For the NPIs to be effective, one of the critical conditions is the public's active participation and compliance. Since the lockdown of Wuhan City on January 23, 2019, China was the first country to introduce NPIs with strict measures such as the lockdown of cities and counties; compulsory mask wearing; isolation of suspicious cases; screening and contact tracing; quarantining people from high risk areas for 14 days; as well as promoting handwashing, proper coughing habits, social distancing, and self-isolation. However, there was no published evidence on relationships between a COVID-19 infection, the Chinese public risk perception, information source, knowledge, attitude, and personal NPIs during the middle to the end of the epidemic.

Between February 22, 2020, to March 5, 2020 (the late period of the COVID-19 epidemic in China) [2], we conducted an online cross-sectional survey of Chinese residents to understand risk perceptions, information source, knowledge, attitude, and practice of the Chinese public after the COVID-19 outbreak; explore the determinants associated with the key personal NPIs (ie, handwashing, proper coughing habits, social distancing, mask wearing); estimate the risks between the COVID-19 infection and the four NPIs; and understand potential risk compensating effects among the four NPIs in relation to the COVID-19 infection (eg, can wearing a mask further reduce the risk of infection among those who do and do not practice the other three NPIs?).

Methods

Study Sample

We conducted an online survey between February 22, 2020 (with total confirmed cases of 77,000 and daily cases of 1500), to March 5, 2020 (with a confirmed total case of 80,500 and daily cases of 151). Given that the whole of China was in lockdown during this period, it was almost impossible to conduct a random sample survey. We chose to conduct our study through the Chinese social media app Wechat (similar to "WhatsApp") and Weibo (similar to "Twitter"). We adopted a snowballing sampling methodology through three social networks: (1) students and staff at Tongji Medical College and Chongqing Medical University; (2) Wanzhou District Centre for Disease Prevention and Control, Chongqing Municipality; and (3) the study team. The inclusion criteria were Chinese citizens who were currently living in Mainland China during the study period, having a mobile phone or computer, and willing to answer all questions. The exclusion criteria were those who did not consent to participate, those who did not answer all the questions, questionnaires completed in less than 2 minutes, and repeated questionnaires from the same Internet Protocol address. During the study period, the survey web page was browsed 21,673 times with a total of 8431 questionnaires returned. After excluding those illegible questionnaires and those who were younger than 18 years, the final study sample was 8158.

The Ethics Committee of Chongqing Medical University approved our study protocol. There was an introduction document before the study questionnaire that provided the respondents with the background, aims, and estimated time (10 minutes) for completing the survey. Respondents were asked for their agreement to participate in the study and to answer the questions faithfully, and were assured confidentiality, anonymity, and that no individual data would be disclosed. After the confirmation of their willingness to participate in the study voluntarily, the participants were directed to complete the online questionnaire. We plan to disseminate the results to study participants whenever appropriate.

The Roles of the Funding Bodies

The funding bodies played no role in the study design, conduct, analysis, interpretation, and the decision to publish the results.

Measures

A multidisciplinary team of 11 experts were involved in the development of the survey instrument. The team included two epidemiologists, two sociologists, one administrative specialist, one statistician, one psychologist, and four postgraduate students. The research team initially had serial meetings to decide the research aims, methodology, as well as the responsibilities of all team members, including their roles in the literature review and seeking ethics approval. The team then developed the survey instrument that included title, number, and content of all sections as well as every question within a

section through an iterative process, which resulted in four major revisions of the first draft. The team also conducted two pilot tests. A group of 15 postgraduate students and staff members of the School of Public Health and Management of Chongqing Medical University participated in two testing pilots of the survey instrument (one test for the earlier version and one test for the near final version). As part of the pilot test, the group were asked a set of questions and then participated in an interview with one of the study team members. The interviews focused on both the survey questionnaires and the user-friendliness of its application in Webchat and Weibo. The discussions around the questionnaire were focused on the clarity of the questions, the readability of the questionnaire, the length of the survey, the overlapping and volume of questions among different sections, and the logic clarity between linked questions. The issues raised during the pilot testing were discussed during the team meetings, and necessary revisions were made. The final instrument included six sections and 79 questions in total, including (1) demographic information (16 questions), (2) knowledge and preventive behaviors related to COVID-19 (2 questions), (3) health status and related health behaviors (32 questions), (4) information source during the COVID-19 epidemic (7 questions), (5) perception and preparedness related to COVID-19 (15 questions), (6) satisfaction with the government's performance in containing the COVID-19 epidemic (6 questions), and an open-ended question asking respondents to share their most important thoughts on the situation. A copy of the final questionnaire is attached (Multimedia Appendices 1 and 2).

The final questions included in this study were personal and family demographics including age, gender, location of residence, education, occupation, family monthly income, smoking habits during the last month (with over 100 cigarettes smoked over the lifetime), drinking alcohol during the last month, height and weight, been infected with COVID-19, marital status, if one of the family members is a health professional, the severity of the community infection where the respondent was living, and if one of the family members was part of local community efforts against COVID-19; perceived risk, attitude, information source, knowledge, and the four NPIs, and if the respondent had repeatedly used a mask; and self-isolation including if the respondents had a Chinese New

Year party (2 days: January 24-25, 2020) with invited guests, the main reason a family member stayed home the longest, the main reason a family member went out often, the approach taken by the respondent when they went out (ie, shorten the time to avoid infection, as usual, stay longer given the restrictions, or uncertainty).

Statistical Analysis

Frequencies of demographic, perceived risk, knowledge, attitude, and the four NPIs as well as self-isolation behaviors were described. The risks between the COVID-19 infection and four binary NPIs were tested using Fisher exact tests. The absolute risk difference, risk ratio (RR), and their 95% CIs were also presented. We modeled the four NPIs using logistic regression. The reasons that we also included proper coughing habit as an end point was that the habit may not only potentially reduce other people's risk of developing a COVID-19 infection but may also reduce a person's own risk through enhanced self-protection (eg, turning away from those who did not practice proper coughing habit) or through indirectly influencing other people's coughing behavior as a role model. The predictors included the demographic characteristics, social economics status, family and social environment, perceived risk of situation, attitude (belief), and knowledge on the four NPIs. We explored the risk between a COVID-19 infection and the four NPIs using a similar approach but excluded knowledge, attitude, and risk perception of the four NPIs based on a penalized maximum likelihood function logistic regression [8,9], which provides consistent estimates in situations of sparse event and total separation. The modeling results for the four NPIs separately and combined (model 1-model 5) were compared to the results of the baseline model with only social demographic variables (model 6). RRs, odds ratios (ORs), and their 95% CIs were presented where appropriate. We explored the potential risk compensating effects among the four NPIs through a pairwise NPIs comparison of infection rates and through the comparison of infection rates of wearing a mask across a combination of the other three NPIs. A flowchart of different sample sizes for the modeling of the four NPIs and the COVID-19 infections are presented (Figure 1). The data management and statistical analysis were done through SPSS v25 (IBM Corp) and Stata v16 (StataCorp). *P* values less than .05 were considered as indicative of significance.

Figure 1. A flowchart of sample sizes for the modeling of the four NPIs (Multimedia Appendix 2: eTable 1) and COVID-19 infection (Tables 5 and 6). M: model; NPI: nonpharmaceutical intervention.
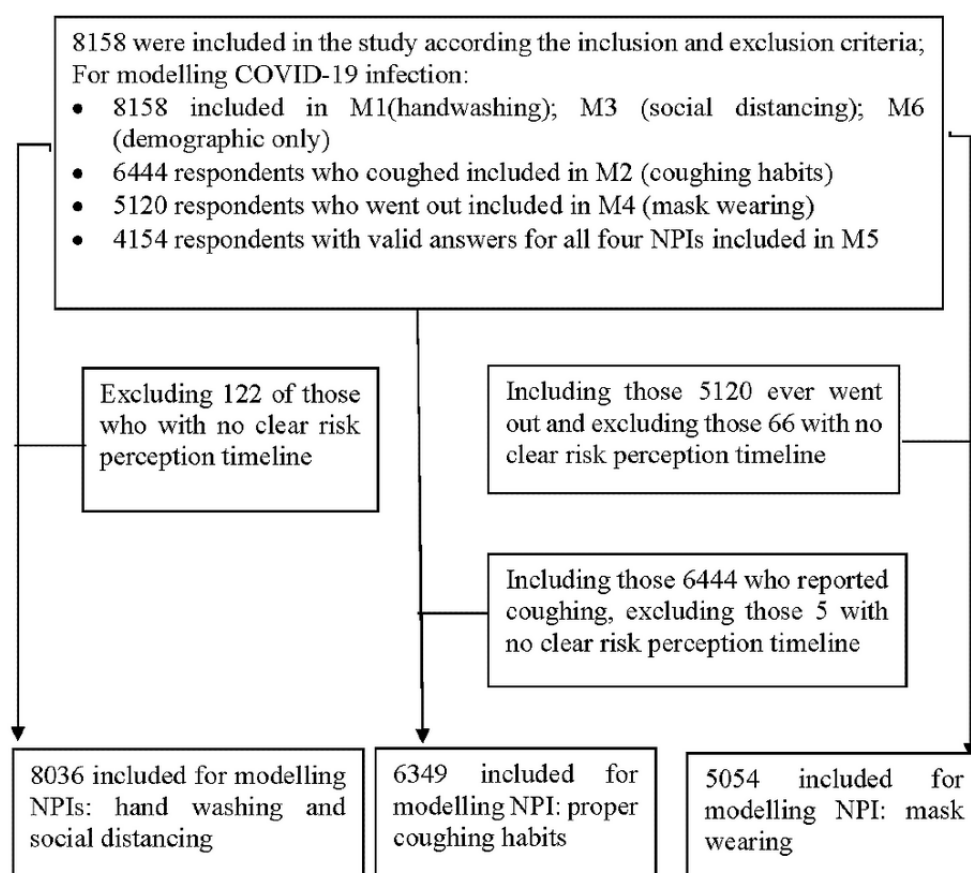
Results

Infection Rate of COVID-19 and Social Demographic of Respondents

In total, 8158 adults were included in the study, and 57 (0.7%) were infected with COVID-19. The respondents were predominantly female (n=5138, 63%), from younger age groups (18-39 years: n=5017, 62%), and living in the city (n=5833, 71.5%). Close to 45% of respondents had undergraduate and above education and close to one-quarter were students (Table 1). The family income from the previous month had a large range (¥0-¥4,000,000; a currency exchange rate of US $1=¥7 is applicable) with close to 20% of them less than ¥2000. Around 13% of respondents smoked and drank alcohol during the previous month. Over half of them had a normal BMI [10]. Close to 72% (n=5854) of respondents had a partner, and 54% (n=4420) were living with one. Close to 37% (n=3001) had a family member who was a health professional, and 35% (n=2835) had a family member who was part of the local community efforts against COVID-19. Over 54% (n=4439) were from the area outside Hubei Province, which had fewer than 100 cases; 42% (n=3400) from the areas outside Hubei Province, which had more than 100 infected cases; and 4% (n=319) from Hubei Province, the epicenter of the Chinese COVID-19 epidemic (Table 1).

Table 1. The demographic characteristics of the study sample.

Characteristics	Participants (N=8158), n (%)
Age groups (years)	
18-39	5017 (61.5)
40-59	2902 (35.6)
≥60	239 (2.9)
Male (vs female)	3020 (37.0)
Currently living in city (vs rural area)	5833 (71.5)
Education	
Primary school	130 (1.6)
High school	2040 (25.0)
Professional college	2331 (28.6)
University/postgraduate	3657 (44.8)
Occupation	
Health professionals	1373 (16.8)
Government payee	1814 (22.2)
Factory workers/managers	1485 (18.2)
Farmers	313 (3.8)
Students	2006 (24.6)
Others	1167 (14.3)
Family monthly income (¥)	
0-1000	607 (7.4)
1001-2000	994 (12.2)
2001-4000	1995 (24.5)
4001-6000	1698 (20.8)
6001-8000	714 (8.8)
8001-10,000	977 (12.0)
10,001-20,000	620 (7.6)
20,001-4,000,000	216 (2.6)
Not sure/unanswered	337 (4.1)
Smoked during the last month (yes)	1087 (13.3)
Drinking during the last month	
Yes	1088 (13.3)
Give up	308 (3.8)
Do not drink	6762 (82.9)
BMI	
Underweight	1574 (19.3)
Normal	4263 (52.3)
Overweight	1264 (15.5)
Obese	133 (1.6)
Not available	924 (11.3)
Infected with COVID-19	57 (0.7)
Do you currently live with your partner?	
Yes	4420 (54.2)

Characteristics	Participants (N=8158), n (%)
No	1434 (17.6)
I do not have a partner	2304 (28.2)
Do you have a health professional family member? (yes)	3001 (36.8)
From the area with community infection	
Hubei Province	319 (3.9)
Outside Hubei Province with ≥100 cases	3400 (41.7)
Other	4439 (54.4)
Is a family member part of local community efforts against COVID-19?	2835 (35.1)

Risk Perception, Information Source, Knowledge, Attitude, the Four NPIs, and Self-Isolation

Of the 8158 respondents, close to 7% (n=561) of respondents were aware of the situation's seriousness on January 11, 2020, when the first COVID-19–related patient death was announced by the Wuhan Health Commission; 39% (n=3182) on January 20, 2020, with the announcement of COVID-19 transmission among humans; 29% (n=2354) on January 23, 2020, with the lockdown of Wuhan City; and 24% (n=1562) after January 24, 2020, due to the activation of the level 1 public emergency responding scheme by local governments as well as strict measures and lockdown of neighborhood or villages [11]. Only 9 (0.1%) respondents did not think it was serious at the time (Table 2). Overwhelmingly, the majority (n=8094, 99.2%) strongly agreed with the position that the fight against COVID-19 was everyone's responsibility (Table 3). Close to

97% perceived government websites, apps, and the public media as the most authoritative sources of information; 90% (n=7396) felt that government websites, apps, and the public media were also the most involved source of information; 99.6% (n=8129) of respondents knew why and how to wash hands properly during the COVID-19 outbreak period; 97.2% (n=7927) were aware of the proper procedures when coughing (turning away from people and covering mouth and nose when coughing and washing hands afterwards); 97.8% (n=7975) knew the right way of practicing social distancing (ie, keeping social distance more than 1 meter and avoiding close contact with those who had a fever or cough); and 99.9% (n=8146) knew why and how to wear a mask. The overwhelming majority also reported that they translated this knowledge into practice: 96.8% (n=7895) when washing hands, 93.1% (n=5997/6444) when coughing, 87.1% (n=7104) when social distancing, and 97.9% (5012/5120) when wearing a mask (Table 3).

Table 2. When participants felt the situation was serious.

Events	Participants (N=8158), n
Announcement of first death of patients with COVID-19 on January 13, 2020	561
Announcement of COVID-19 transmission among humans on January 19, 2020	3182
Lockdown of Wuhan City on January 23, 2020	2354
Activation of the public emergency responding scheme by local government on January 24, 2020	1562
Strict lockdown of neighborhood after January 24, 2020	377
Other critical events	113
Never felt it was serious	9

Our multivariable models found that the knowledge of the relevant NPIs was one of the strongest predictors of that behavior (OR 22.6 for handwashing, OR 4.26 for social distancing; all *P*<.001), and positive knowledge level associated with a proper coughing habit and mask wearing were at 100% and were excluded from the models (Multimedia Appendix 3, first eTable). The belief that the task of fighting against COVID-19 is everyone's responsibility was positively associated with handwashing (OR 5.59, *P*<.001), social distancing (OR 3.76, *P*<.001), and mask wearing (OR 26.89, *P*<.001). Those who perceived the seriousness of the situation before the lockdown of Wuhan City were more likely to practice handwashing (OR 1.43, 95% CI 1.08-1.90; *P*=.01) and a proper coughing habit (OR 1.54, 95% CI 1.24-1.92; *P*<.001). Those who had a family member involved in the local community

efforts against COVID-19 and people from outside the Hubei Province were more likely associated with positive NPIs. In comparison to those who were currently living with a partner, those who did not have partner were less likely to practice handwashing (OR 0.57, 95% CI 0.35-0.95; *P*=.03) and proper coughing habits (OR 0.59, 95% CI 0.41-0.83; *P*=.003) but were more likely to practice social distancing (OR 1.64; *P*<.001). The respondents who had a normal body weight were more likely to practice social distancing (OR 1.21, 95% CI 1.01-1.46; *P*=.04) than those who were underweight. Nonsmokers were more likely to practice social distancing (OR 1.39, 95% CI 1.14-1.71; *P*=.001) than smokers. Family income, education, occupation, residential area, sex, and age groups demonstrated differential impacts on different NPIs (Multimedia Appendix 3, first eTable).

Approximately 23% of respondents reported they had a Chinese New Year party with invited guests (Multimedia Appendix 3, second eTable). Out of 8158 respondents, the major reasons for the family member who stayed longest at home included complying with the call from the government (n=5354, 65.6%), self- or compulsory isolation (n=1107, 13.6%), fear of the virus (n=469, 5.7%), the focus on family protection (n=310, 3.8%), and no mask (n=242, 3.0%). The main reasons for going out, for 5120 respondents, were shopping (n=2073, 40.5%), partaking

in work related to controlling COVID-19 (n=1643, 32.1%), usual employment (n=1098, 21.4%), going for a walk (n=131, 2.6%), receiving delivery (n=38, 0.7%), and socializing and dinner parties (n=17, 0.3%). Over 74% shortened the time to avoid infection when they were out; close to 20% acted in the usual way; and 1% (n=50) stayed longer than usual, given the restrictions and difficulties to leave home (Multimedia Appendix 3 second eTable).

Table 3. Attitude, source of information, knowledge, and the four nonpharmaceutical interventions.

Item	Participant (N=8158), n (%)
Agree with that the fight with COVID-19 is everyone's job	8094 (99.2)
Perceived most authoritative source of information	
Government websites/app/public media	7902 (96.9)
Weibo/Webchat friends	128 (1.6)
QQ/webchat groups	64 (0.8)
Family/friends	40 (0.5)
Other	24 (0.3)
Perceived most interested source of information	
Government websites/app/public media	7396 (90.7)
Weibo/Webchat friends	470 (5.8)
QQ/webchat groups	218 (2.7)
Family/friends	60 (0.7)
Other	14 (0.2)
Know why and how to wash hands	8129 (99.6)
Know the proper habit when coughing	7927 (97.2)
Know why and how to practice social distancing	7975 (97.8)
Know why and how to wear a mask in public	8146 (99.9)
Washing hands (yes)	7895 (96.8)
Acting in proper habit when coughing[a]	5997 (93.1)
Practicing social distancing	7104 (87.1)
Wearing a mask when going out[b]	5012 (97.9)
Repeated use of a mask[b]	5117 (99.9)

[a]Only 6444 included as 1714 person reported that they did not cough during the last month.

[b]Only the 5120 respondents that went out during the period after the outbreak were included.

Risk Association Between a COVID-19 Infection and Handwashing, Coughing Habits, Social Distancing, and Mask Wearing

The distribution of the COVID-19 infection across demographic and social economic status variables are presented (Table 4). The bivariate analyses between individual NPIs and the COVID-19 infection showed that there was a significantly increased risk of COVID-19 infection (Table 5) for those who did not wash their hands (2.28% vs 0.65%; RR 3.53, 95% CI 1.53-8.15; P=.009), who did not practice proper coughing (1.79% vs 0.73%; RR 2.44, 95% CI 1.15-5.15; P=.03), who did not practice social distancing (1.52% vs 0.58%; RR 2.63, 95%

CI 1.48-4.67; P=.002), and who did not wear a mask (7.41% vs 0.6%; RR 12.38, 95% CI 5.81-26.36; P<.001). The adjusted ORs were 4.67 (95% CI 1.86-11.74; P=.001) for not washing hands, 2.78 (95% CI 1.22-6.33; P=.02) for not practicing proper coughing, 2.13 (95% CI 1.17-3.85; P=.01) for not practicing social distancing, and 11.03 (95% CI 4.53-26.84; P<.001) for not wearing a mask (Tables 6 and 7, model 1-model 4). The model that adjusted all four NPIs plus social demographic variables (Table 7, model 5) showed that not wearing a mask was the only significant predictor of infection (OR 7.20, 95% CI 2.24-23.11; P<.001). In comparison with those who were only primary school educated, those with a high school qualification showed they were less likely to be infected (OR 0.12, 95% CI 0.05-0.31; P<.001). This was similar to those who

had professional college qualifications (OR 0.10, 95% CI 0.03-0.29; *P*<.001) or university degrees (OR 0.15, 95% CI 0.05-0.46; *P*=.001; Table 7, model 6). Nonsmokers were less likely to be infected than smokers (OR 0.40, 95% CI 0.20-0.80; *P*=.01), and those with a monthly family income of ¥8001-¥10,000 were less likely to be infected than those having a monthly family income of less than ¥1000 (OR 0.20, 95% CI 0.05-0.88; *P*=.03; Table 7, model 6). Consisting of the effective sizes of the four NPIs from both bivariable and multivariable

analyses, the areas under the receiver operating characteristics curve (AUROCs) were 0.749, 0.769, 0.749, and 0.825 for handwashing, coughing habits, social distancing, and mask wearing, respectively (Tables 6 and 7, model 1-model 4). The AUROC values demonstrated a fair predictive power for the handwashing, coughing habits, and social distancing models but good predictive power of the mask wearing multivariable model.

Table 4. The rate of COVID-19 infection (57/8158, 0.7%) across social demographic variables.

Variables	Participants, n	Participants infected with COVID-19, n (%)	P value[a]
Age groups (years)			.14
18-39	5017	29 (0.58)	
40-59	2902	25 (0.86)	
≥60	239	3 (1.26)	
Sex			.008
Male	3020	31 (1.03)	
Female	5138	26 (0.51)	
The area that you are currently living			.46
Rural	2325	19 (0.82)	
City	5833	38 (0.65)	
Education			<.001
Primary school	130	8 (6.15)	
High school	2040	14 (0.69)	
Professional college	2331	13 (0.56)	
University/postgraduate	3657	22 (0.60)	
Occupation			.17
Health professionals	1373	9 (0.66)	
Government payee	1814	14 (0.77)	
Factory workers/managers	1485	14 (0.94)	
Farmers	313	4 (1.28)	
Students	2006	7 (0.35)	
Others	1167	9 (0.77)	
Income (¥)			.20
0-1000	607	8 (1.32)	
1001-2000	994	7 (0.70)	
2001-4000	1995	16 (0.80)	
4001-6000	1698	13 (0.77)	
6001-8000	714	4 (0.56)	
8001-10,000	977	2 (0.20)	
10,001-20,000	620	5 (0.81)	
20,001-4,000,000	216	2 (0.93)	
Not sure/unanswered	337	0 (0.00)	
Smoked during the last month			<.001
Yes	1087	20 (1.84)	
No	7071	37 (0.52)	
Drank alcohol during the last month			.002
Yes	1088	12 (1.10)	
Gave up	308	7 (2.27)	
Do not drink	6762	38 (0.56)	
BMI			.85
Underweight	1574	12 (0.76)	
Normal	4263	29 (0.68)	

Variables	Participants, n	Participants infected with COVID-19, n (%)	P value[a]
Overweight	1264	7 (0.55)	
Obesity	133	1 (0.75)	
Not available	924	8 (0.87)	
Currently live with your partner?			.01
Yes	4420	36 (0.81)	
No	1434	14 (0.98)	
I do not have a partner	2304	7 (0.30)	
Family member who is a health professional			.17
Yes	3001	26 (0.87)	
No	5157	31 (0.60)	
From the area with community infection			.60
Hubei Province	319	3 (0.94)	
Outside Hubei province with ≥100 cases	3400	21 (0.62)	
Other	4439	33 (0.74)	
Family member who is part of local community efforts against COVID-19			.17
Yes	2865	25 (0.87)	
No	5293	32 (0.60)	

[a]P values from Fisher exact test.

Table 5. The association between COVID-19 infection and four nonpharmaceutical interventions.

Nonpharmaceutical interventions	Risk of COVID-19 infection, n (%)	Risk difference (95% CI), %	Risk ratio (95% CI)	P value[a]
Washing hands (N=8158)		1.63 (–0.18 to 3.45)	3.53 (1.53 to 8.15)	.009
No (n=263)	6 (2.28)			
Yes (n=7895)	51 (0.65)			
Acting properly when coughing (n=6444)		1.06 (–0.19 to 2.3)	2.44 (1.15 to 5.15)	.03
No (n=447)	8 (1.79)			
Yes (n=5997)	44 (0.73)			
Practicing social distancing (N=8158)		0.94 (0.18 to 1.7)	2.63 (1.48 to 4.67)	.002
No (n=1054)	16 (1.52)			
Yes (n=7104)	41 (0.58)			
Wearing a mask outdoors (n=5120)		6.81 (1.87 to 11.75)	12.38 (5.81 to 26.36)	<.001
No (n=108)	8 (7.41)			
Yes (n=5012)	30 (0.60)			

[a]P values from Fisher exact test.

Table 6. Penalized logistic regression model results for COVID-19 infection and handwashing, coughing habits, social distancing, and mask wearing (models 1-3).[a]

Variables	Infected (model 1; N=8158), OR[b] (950% CI)	Infected (model 2; n=6444), OR (95% CI)	Infected (model 3; N=8158), OR (95% CI)
Age groups (years), reference: 18-39 years			
40-59	0.97 (0.51-1.85)	0.88 (0.45-1.73)	0.93 (0.49-1.78)
≥60	1.55 (0.46-5.25)	1.68 (0.49-5.83)	1.71 (0.51-5.75)
Sex (female vs male)	0.76 (0.39-1.47)	0.72 (0.36-1.44)	0.76 (0.39-1.48)
Living in city (urban vs rural)	0.94 (0.50-1.77)	0.81 (0.42-1.56)	0.91 (0.48-1.71)
Education, reference: primary school			
High school	0.13 (0.05-0.33)***	0.12 (0.04-0.34)***	0.12 (0.05-0.31)***
Professional college	0.10 (0.03-0.30)***	0.10 (0.03-0.31)***	0.10 (0.03-0.29)***
University/postgraduate	0.16 (0.05-0.48)**	0.14 (0.04-0.46)**	0.16 (0.05-0.48)**
Occupation, reference: health professionals			
Government payee	1.51 (0.63-3.67)	1.67 (0.65-4.28)	1.54 (0.63-3.73)
Factory workers/managers	1.82 (0.66-5.00)	2.12 (0.74-6.09)	1.87 (0.68-5.17)
Farmers	1.02 (0.25-4.14)	0.95 (0.21-4.39)	0.99 (0.24-4.13)
Students	0.97 (0.26-3.69)	0.90 (0.22-3.70)	1.10 (0.29-4.17)
Others	1.63 (0.51-5.19)	1.79 (0.52-6.09)	1.73 (0.54-5.55)
Family monthly income (¥), reference: ¥0-¥1000			
1001-2000	0.70 (0.25-1.96)	0.58 (0.20-1.75)	0.61 (0.22-1.68)
2001-4000	0.79 (0.32-1.93)	0.69 (0.27-1.74)	0.73 (0.30-1.77)
4001-6000	0.68 (0.27-1.71)	0.62 (0.24-1.62)	0.63 (0.25-1.57)
6001-8000	0.54 (0.16-1.83)	0.56 (0.16-1.93)	0.50 (0.15-1.69)
8001-10,000	0.21 (0.05-0.93)*	0.13 (0.02-0.76)*	0.20 (0.05-0.88)*
10,001-20,000	0.67 (0.20-2.22)	0.62 (0.19-2.07)	0.63 (0.20-2.06)
20,001-4,000,000	0.54 (0.12-2.51)	0.53 (0.11-2.48)	0.55 (0.12-2.45)
Not sure/unanswered	0.15 (0.01-2.63)	0.15 (0.01-2.62)	0.13 (0.01-2.32)
Smoked during the last month (no vs yes)	0.38 (0.19-0.76)**	0.38 (0.18-0.77)**	0.42 (0.21-0.84)*
Drank alcohol during the last month, reference: yes			
Gave up	1.96 (0.76-5.04)	2.07 (0.79-5.44)	1.94 (0.75-4.97)
Do not drink	0.88 (0.42-1.87)	0.87 (0.40-1.89)	0.84 (0.40-1.78)
BMI, reference: underweight			
Normal	0.80 (0.41-1.59)	0.76 (0.37-1.57)	0.80 (0.41-1.59)
Overweight	0.49 (0.19-1.25)	0.55 (0.21-1.43)	0.49 (0.19-1.26)
Obese	0.88 (0.16-4.94)	1.06 (0.19-6.06)	0.94 (0.17-5.28)
Not available	1.15 (0.46-2.86)	1.17 (0.45-3.07)	1.11 (0.45-2.78)
Currently live with your partner, reference: yes			
No	1.34 (0.69-2.63)	1.36 (0.67-2.77)	1.37 (0.70-2.69)
I do not have a partner	0.50 (0.19-1.34)	0.71 (0.26-1.90)	0.55 (0.21-1.49)
Family member who is health professional (yes vs no)	0.53 (0.26-1.08)	0.59 (0.28-1.23)	0.50 (0.24-1.01)
Living in the area with a community infection, reference: Hubei Province			
Outside Hubei Province with ≥100 cases	0.46 (0.14-1.53)	0.46 (0.13-1.61)	0.49 (0.14-1.65)
Other	0.74 (0.23-2.36)	0.82 (0.25-2.63)	0.81 (0.25-2.62)

Variables	Infected (model 1; N=8158), OR[b] (950% CI)	Infected (model 2; n=6444), OR (95% CI)	Infected (model 3; N=8158), OR (95% CI)
Family member who is part of the local community efforts against COVID-19 (no vs yes)	0.70 (0.37-1.30)	0.64 (0.34-1.22)	0.76 (0.41-1.42)
Hand washing (no vs yes)	4.67 (1.86-11.74)**	N/A[c]	N/A
Proper coughing habit (no vs yes)	N/A	2.78 (1.22-6.33)*	N/A
Social distancing (no vs yes)	N/A	N/A	2.13 (1.17-3.85)*
Mask wearing (no vs yes)	N/A	N/A	N/A
Constant	0.67 (0.09-4.81)	0.90 (0.11-6.99)	0.60 (0.08-4.42)
AUROC[d]	0.749	0.769	0.749

[a]Models 1-5 are models adjusting for social demographic variables plus handwashing, coughing habits, social distancing, mask wearing, and all four nonpharmaceutical interventions together, respectively. Model 6 is adjusting for social demographic variables only. For each model, the convergency of penalized maximum likelihood function was monitored, and coefficients and their 95% CIs were examined. Each model converged normally in a short period and no irregularity of coefficients and standard errors were identified.

[b]OR: odds ratio.

[c]N/A: not applicable.

[d]AUROC: area under the receiver operating characteristic curve.

*$P<.05$, **$P<.01$, ***$P<.001$

Table 7. Penalized logistic regression model results for COVID-19 infection and handwashing, coughing habits, social distancing, and mask wearing.[a]

Variables	Infected (model 4; n=5120), OR[b] (95% CI)	Infected (model 5; n=4154), OR (95% CI)	Infected (model 5; N=8,158), OR (95% CI)
Age groups (years), reference: 18-39 years			
40-59	1.07 (0.47-2.42)	1.17 (0.49-2.81)	0.95 (0.50-1.81)
≥60	3.16 (0.82-12.11)	3.79 (0.96-14.94)	1.68 (0.50-5.71)
Sex (female vs male)	0.65 (0.29-1.49)	0.69 (0.29-1.66)	0.73 (0.38-1.42)
Living in city (urban vs rural)	0.85 (0.37-1.95)	0.75 (0.32-1.77)	0.89 (0.47-1.68)
Education, reference: primary school			
High school	0.09 (0.03-0.30)***	0.11 (0.03-0.43)**	0.12 (0.05-0.31)***
Professional college	0.09 (0.03-0.34)***	0.11 (0.03-0.47)**	0.10 (0.03-0.29)***
University/postgraduate	0.08 (0.02-0.33)***	0.10 (0.02-0.47)**	0.15 (0.05-0.46)***
Occupation, reference: health professionals			
Government payee	1.74 (0.60-5.06)	1.59 (0.52-4.85)	1.57 (0.65-3.81)
Factory workers/managers	1.97 (0.56-6.89)	1.80 (0.49-6.55)	1.95 (0.71-5.33)
Farmers	1.35 (0.25-7.27)	0.76 (0.12-4.97)	1.08 (0.26-4.43)
Students	1.00 (0.19-5.36)	0.69 (0.12-4.03)	1.18 (0.31-4.43)
Others	2.02 (0.49-8.27)	2.17 (0.52-9.04)	1.72 (0.54-5.51)
Family monthly income (¥), reference: ¥0-¥1000			
1001-2000	0.42 (0.11-1.57)	0.58 (0.14-2.38)	0.62 (0.23-1.72)
2001-4000	0.77 (0.26-2.28)	0.90 (0.28-2.92)	0.73 (0.30-1.78)
4001-6000	0.47 (0.14-1.52)	0.64 (0.18-2.25)	0.63 (0.25-1.58)
6001-8000	0.56 (0.13-2.37)	0.73 (0.17-3.25)	0.50 (0.15-1.68)
8001-10,000	0.31 (0.06-1.49)	0.24 (0.03-1.62)	0.20 (0.05-0.88)*
10,001-20,000	0.43 (0.10-1.87)	0.53 (0.11-2.49)	0.62 (0.19-2.03)
20,001-4,000,000	0.24 (0.02-2.61)	0.21 (0.01-3.85)	0.60 (0.14-2.64)
Not sure/unanswered	0.11 (0.00-2.75)	0.26 (0.01-4.97)	0.13 (0.01-2.36)
Smoked during the last month (no vs yes)	0.67 (0.29-1.58)	0.65 (0.27-1.57)	0.40 (0.20-0.80)**
Drank alcohol during the last month, reference: yes			
Gave up	1.29 (0.43-3.87)	1.34 (0.43-4.20)	2.00 (0.78-5.10)
Do not drink	0.48 (0.20-1.13)	0.47 (0.19-1.15)	0.85 (0.40-1.80)
BMI, reference: underweight			
Normal	0.56 (0.25-1.25)	0.57 (0.24-1.33)	0.79 (0.40-1.57)
Overweight	0.26 (0.07-0.90)*	0.28 (0.08-0.99)*	0.48 (0.19-1.25)
Obese	0.31 (0.02-5.81)	0.38 (0.02-7.45)	0.88 (0.16-4.97)
Not available	1.25 (0.46-3.44)	1.17 (0.40-3.46)	1.13 (0.45-2.82)
Currently living with a partner, reference: yes			
No	1.62 (0.69-3.82)	1.28 (0.50-3.26)	1.38 (0.70-2.69)
I don't have a partner	0.98 (0.30-3.20)	1.19 (0.36-3.91)	0.53 (0.20-1.41)
Family member who is a health professional (yes vs no)	0.51 (0.22-1.19)	0.61 (0.25-1.48)	0.50 (0.24-1.01)
Living in an area with community infection, reference: Hubei Province			
Outside Hubei Province with ≥100 cases	0.24 (0.06-0.95)*	0.25 (0.06-1.09)	0.45 (0.13-1.51)
Other	0.68 (0.19-2.50)	0.74 (0.20-2.81)	0.78 (0.25-2.47)

XSL·FO
RenderX

Variables	Infected (model 4; n=5120), OR[b] (95% CI)	Infected (model 5; n=4154), OR (95% CI)	Infected (model 5; N=8,158), OR (95% CI)
Family member who is part of the local community efforts against COVID-19 (no vs yes)	0.51 (0.24-1.09)	0.46 (0.21-1.04)	0.74 (0.40-1.37)
Handwashing (no vs yes)	N/A[c]	1.82 (0.40-8.32)	N/A
Proper coughing habit (no vs yes)	N/A	1.88 (0.60-5.94)	N/A
Social distancing (no vs yes)	N/A	1.07 (0.46-2.46)	N/A
Mask wearing (no vs yes)	11.03 (4.53-26.84)***	7.20 (2.24-23.11)***	N/A
Constant	2.04 (0.19-21.47)	1.64 (0.13-20.42)	0.78 (0.11-5.62)
AUROC[d]	0.825	0.823	0.748

[a]Models 1-5 are models adjusting for social demographic variables plus handwashing, coughing habits, social distancing, mask wearing, and all four nonpharmaceutical interventions together, respectively; model 6 is adjusting for social demographic variables only. For each model, the convergency of penalized maximum likelihood function was monitored, and coefficients and their 95% CIs were examined. Each model converged normally in a short time and no irregularity of coefficients and standard errors was identified.

[b]OR: odds ratio.

[c]N/A: not applicable.

[d]AUROC: area under the receiver operating characteristic curve.

*$P<.05$, **$P<.01$, ***$P<.001$

Potential Risk Compensating Effects Among the Four NPIs Against a COVID-19 Infection

The pairwise distributions of the COVID-19 infection rate among the four NPIs are presented (Table 8). Wearing a mask (vs not) was associated with a significantly reduced risk of COVID-19 infection among those who practiced handwashing (0.6% vs 5.3%; RR 0.11; $P<.001$), proper coughing (0.7% vs 3.9%; RR 0.18; $P=.02$), and social distancing (0.5% vs 16.7%; RR 0.03; $P=.002$). Handwashing showed a trend toward a further reduced risk of infection for those who did not practice social distancing (RR 0.25; $P=.05$). Among those who did not practice social distancing, those who had proper coughing habits were associated with a reduced risk of infection compared to those who did not have proper coughing habits (1.3% vs 4.4%; RR 0.29; $P=.048$). The added potential protection effect of mask wearing on different combinations of the other three NPIs are presented in Figure 2. For those who did practice all three NPIs (ie, handwashing, proper coughing, and social distancing), wearing a mask was associated with a significantly reduced risk of infection compared to those who did not (0.6% vs 16.7%; $P=.04$). Similarly, for those who did not practice all other three NPIs, wearing a mask was also associated with a significantly reduced risk of infection compared to not wearing a mask (Figure 2).

Table 8. The COVID-19 infection rates (%) and potential pairwise risk compensating effect among the four nonpharmaceutical interventions with RR and *P* values from exact tests.

Nonpharmaceutical interventions	Participants, n (%)	RR[a] (95% CI)	*P* value
Handwashing (no)		0.03 (0.003-0.26)	.002
Mask wearing (no)	13 (23.1)		
Mask wearing (yes)	147 (0.7)		
Handwashing (yes)		0.11 (0.04-0.29)	<.001
Mask wearing (no)	95 (5.3)		
Mask wearing (yes)	4865 (0.6)		
Proper coughing (no)		0.02(0.004-0.08)	<.001
Mask wearing (no)	11 (36.4)		
Mask wearing (yes)	318 (0.6)		
Proper coughing (yes)		0.18 (0.05-0.57)	.02
Mask wearing (no)	76 (3.9)		
Mask wearing (yes)	3749 (0.7)		
Social distancing (no)		0.17 (0.06-0.45)	.002
Mask wearing (no)	96 (6.3)		
Mask wearing (yes)	958 (1.0)		
Social distancing (yes)		0.03 (0.01-0.11)	.002
Mask wearing (no)	12 (16.7)		
Mask wearing (yes)	4054 (0.5)		
Proper coughing (no)		0.33 (0.08-1.35)	.13
Handwashing (no)	74 (4.1)		
Handwashing (yes)	373 (1.3)		
Proper coughing (yes)		0.42 (0.10-1.72)	.21
Handwashing (no)	118 (1.7)		
Handwashing (yes)	5879 (0.7)		
Social distancing (no)		0.25 (0.73-0.86)	.05
Handwashing (no)	58 (5.2)		
Handwashing (yes)	996 (1.3)		
Social distancing (yes)		0.38 (0.12-1.21)	.11
Handwashing (no)	205 (1.5)		
Handwashing (yes)	6899 (0.6)		
Social distancing (no)		0.29(0.09-0.91)	.048
Proper coughing (no)	90 (4.4)		
Proper coughing (yes)	776 (1.3)		
Social distancing (yes)		0.58 (0.21-1.63)	.30
Proper coughing (no)	357 (1.1)		
Proper coughing (yes)	5221 (0.7)		

[a]RR: risk ratio.

Figure 2. The COVID-19 infection rates between mask wearing and the combinations of the other three nonpharmaceutical interventions: H, C, and S. *P* values from Fisher exact test. C: proper coughing habit; H: handwashing; S: social distancing.
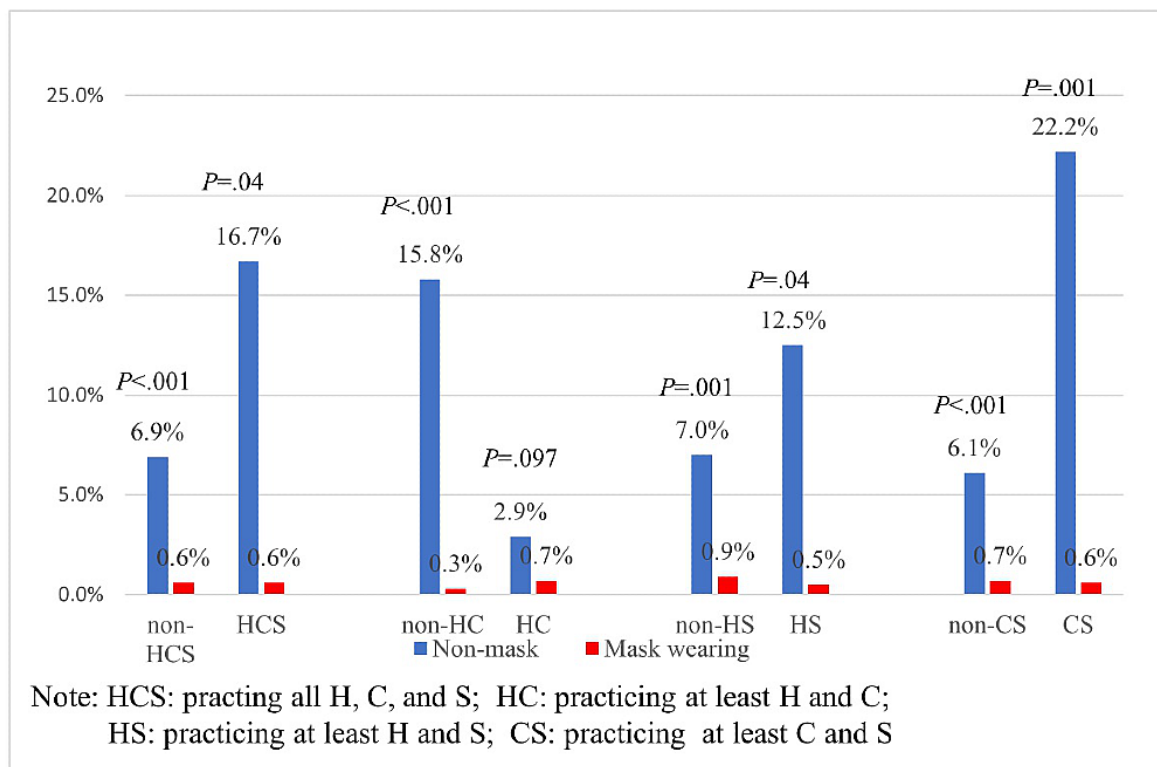
Discussion

To the best of our knowledge, our study is the first comprehensive report on the COVID-19 infection rate, perceived risk, knowledge, attitude, the four NPIs, as well as the self-isolation of a nationwide adult sample amid the late period of the COVID-19 epidemic in China. We found that most respondents were aware of the outbreak's seriousness at different time periods and believed that it was everyone's responsibility to fight the spread of COVID-19. The positive attitude, earlier risk perception, and relevant knowledge were among the strongest predictors of handwashing, proper coughing habits, social distancing, and mask wearing. Different social demographic factors also contributed to different NPIs. Those who were only primary school educated, had little family income, and were a smoker were associated with an increased risk of COVID-19 infection. Mask wearing, among the four practices, was the most important protection factor against a COVID-19 infection with an added preventive effect among those who practiced all or part of the other three NPIs.

Our findings of high knowledge levels among the Chinese public were consistent with the results of a previous study [12]. However, the previous online survey study was conducted at a much earlier stage (ie, January 27 to February 1, 2020) with a smaller sample, and over half of the respondents were from Hubei Province. In addition, it did not include the knowledge and behaviors of handwashing, proper coughing habit, social distancing, and self-isolation (but mask wearing was included). The widespread use of mobile phones, the internet, and social media apps such as Wechat (with an estimated 1.1 billion registered accounts in China in 2019 [13]) have significantly

increased the speed and scope of information transmission in China and may be instrumental in the dissemination of COVID-19–related knowledge and information. This was also helped by the fact that over 90% of respondents believed that the government websites, public media, and apps were the most authoritative and involved sources of information for COVID-19. Our results suggested that respondents had an extremely positive attitude that fighting against COVID-19 was everyone's responsibility, and its association with positive NPIs highlighted the success of a nationwide campaign in instilling the concept that everyone can, and should, make necessary contributions toward the fight against COVID-19.

Our study's finding that the early perceived seriousness of the situation was a strong predicting factor for the use of NPIs reinforces the importance of transparency and timely dissemination of critical information regarding the COVID-19 pandemic. Our study shows that around 22% of respondents had a party during Chinese New Year (January 24-25, 2020), which is an important Chinese tradition. This period was immediately after the lockdown of Wuhan City (January 23), where there was confusion for those who lived outside the Hubei Province where there was no social distancing and self-isolation in place at the time. Such gatherings could have been avoided if the public had been equipped with real-time knowledge of the danger and seriousness of the situation. Evidence also supported the concept that the early swift NPIs by governments and the education of the public regarding the seriousness of the situation were critical in slowing the spread and flattening the curve [14]. Our study demonstrates that family influences (particularly those with health professionals and with someone being part of a community team fighting against COVID-19) could have a significant positive impact on individual's

behaviors. Respondents with various demographic characteristics (eg, age groups, sex, marital status, education, occupation, and smoker) had exhibited different NPIs. These findings may provide further opportunities for developing tailored health education campaigns and health policy interventions on segments of the population to maximize the effects of the NPIs. For example, specific policies and education could target smokers and the younger population to encourage certain behaviors such as proper coughing habit and the older population on other behaviors (eg, self-isolation, social distancing, and mask wearing).

The reported COVID-19 infection rate (ie, 57/8158, 0.73%) in our study is higher than the China national infection rate on March 5, 2020 (ie, 0.0056%), that reflected the fact that the disproportional sample were from the higher risk Hubei province. Our study found almost universal acceptance of the importance of mask wearing, and a high proportion (5012/5120, 97.9%) wore a mask in public after the outbreak. Our study found that mask wearing, among the four personal NPIs, was the most important protective measure against a COVID-19 infection. This may have policy implications. The Chinese public accepted the concept of wearing a mask possibly due to factors such as the previous severe acute respiratory syndrome epidemic experience [15], the coordinated nationwide education campaign, the earlier recognition of the existence of asymptomatic virus carriers, the strict measures in reinforcing such a role (eg, in shopping centers and public transportation), and the coordinated efforts in rationing the supply of masks to families over the shortage period. The necessity in wearing a mask in public may be controversial in different countries and agencies [16-20] despite the positive evidence in favoring wearing a mask in a simulated environment [21,22]. It is likely to be an evolving policy option depending on several factors including the availability of masks and fair distribution channels among society. Our findings that wearing a mask had an added preventive effect, even among those who did practice all or part of the other three NPIs, provided contradictive evidence regarding the opinion that the other three NPIs alone were sufficient in preventing the COVID-19 infection. It also did not support the opinion that wearing a mask could even increase the risk of infection through more facial contact. For those countries still in the grip of the pandemic or who are considering reopening their economy, a policy of encouraging or requiring the public to wear a mask may have a positive impact, especially in highly populated areas or in settings where other NPIs are difficult to implement (eg, in a bus, airplane, or shopping center). During the study period, there were still 3% (242/8158) of respondents who reported that *no mask* was the main reason for stopping them from going out, and most respondents had repeatedly used the same mask. Given the likelihood of a surge in demand for masks during an outbreak, public health agencies and related authorities may also need to provide practical and evidence-based guidance on when and how to appropriately reuse a mask. China contributed over half of the global mask manufacturing output before the outbreak but still faced the shortage of masks over the epidemic period [23]. It is important for governments and international agencies to rethink the adequacy of, and better approaches toward, their strategic stockpiles of masks and other personal protection equipment for the current and future pandemics.

Our study has several strengths. First, it was the largest study of its kind to cover the most critical period of the COVID-19 outbreak in China. Second, our study design and analysis was driven by policy needs and included many factors such as demographics, social economic status, family contextual factors, risk perception, knowledge, attitude, and personal practices. Third, the adoption of the internet survey methodology enabled us to complete our study in a critical period and in a cost-effective manner. Our study also has several limitations. First, our study sample had disproportionately more female, well-educated, and less smoker respondents, reflecting a typically young and healthy cohort in similar surveys. Thus, the frequencies of desirable knowledge levels and health behaviors may be overestimated, while less desirable outcomes such as lower family monthly income may be underestimated. However, the modeling results may be less susceptible to these potential biases. Second, our study results were from a particular period of the outbreak, and most of the respondents were from outside Hubei Province. The generalization of the results to other settings and countries may be limited. Third, our study was a cross-sectional population survey, and the association found between the predictors and outcomes should be interpreted with caution, and further intervention studies are needed in confirming our findings. Fourth, despite the relatively large sample size, the total cases of COVID-19 infections were still small so that the relationship between NPIs and a COVID-19 infection should be confirmed by other larger epidemiological studies. Fifth, the potential risk compensating effects of wearing a mask against other NPIs should be considered as being of a hypothesis-generating nature given the potential limitations previously outlined. Sixth, all the information collected in the study was self-reported, which could have potential biases. Common to any observational studies with multiple outcomes and modeled with different effective sample sizes, the interpretations and generalization of the results should be strictly limited to the same setting and be aware of multiple tests risks.

In summary, our study found a high level of risk perception; positive attitude; desirable knowledge; and practices in handwashing, proper coughing habit, social distancing, and mask wearing among a large cohort of Chinese adults. Our study also found that the relevant knowledge, risk perception, and attitude were among the strongest predictors of the four NPIs. Wearing a mask, among the four NPIs, was the predominating protective measure against a COVID-19 infection, with an added preventive effect among those who practiced all or part of the other three NPIs. Our findings of many different predictors on different personal NPIs may also provide the possibility for further tailored health policy interventions. The study also emphasizes the importance, at an international level, of sharing information in a collaborative way to learn from everyone's experiences about what interventions worked well and what were the impact of issues that may have resulted in poor outcomes such as delayed and misinformed actions.

Authors' Contributions

HX, YG, and ZL conceived and designed the study. HX, YG, DZ, BW, XZ, CX, CL, ZT, MC, and ML participated in the acquisition of data. JC, HX, and XZ conceptualized the theoretical and analytical framework, and conducted data management and statistical analysis. ZL and YH provided advice on the methodology. JC, HX, YG, and XZ conducted the literature review, and HX, YG, and XZ provided the first draft of the manuscript. JC produced all the final tables and figures as well as the final draft for review. ZL, YH, and other authors revised the manuscript. All authors read and approved the final manuscript. JC (jackchen@unsw.edu.au) was the submitting and corresponding author up to the publication of the paper. JC is the guarantor of this work, has full access to all the data in the study, and takes responsibility for its integrity and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COVID-19 questionnaire: Chinese version.
[DOCX File , 39 KB - jmir_v22i11e21372_app1.docx]

Multimedia Appendix 2
COVID-19 questionnaire: English version.
[DOCX File , 48 KB - jmir_v22i11e21372_app2.docx]

Multimedia Appendix 3
eTables.
[DOCX File , 34 KB - jmir_v22i11e21372_app3.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. 2020. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
3. Spinney L. Coronavirus vaccine: when will we have one? The Guardian. 2020. URL: https://www.theguardian.com/world/2020/apr/12/when-will-we-have-a-coronavirus-vaccine
4. Lurie N, Saville M, Hatchett R, Halton J. Developing covid-19 vaccines at pandemic speed. N Engl J Med 2020 May 21;382(21):1969-1973. [doi: 10.1056/nejmp2005630]
5. Rubin EJ, Baden LR, Morrissey S. Audio interview: making decisions about covid-19 testing and treatment for your patients. N Engl J Med 2020 Mar 12;382(11):e25. [doi: 10.1056/nejme2004856]
6. Rubin EJ, Baden LR, Morrissey S. Audio interview: new research on possible treatments for covid-19. N Engl J Med 2020 Mar 19;382(12):e30. [doi: 10.1056/nejme2005759]
7. Nonpharmaceutical interventions (NPIs). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/nonpharmaceutical-interventions/index.html [accessed 2020-04-08]
8. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002 Aug 30;21(16):2409-2419. [doi: 10.1002/sim.1047] [Medline: 12210625]
9. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80(1):27-38. [doi: 10.1093/biomet/80.1.27]
10. Body mass index - BMI. World Health Organization. 2020. URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [accessed 2020-03-30]
11. Tay JK, Khoo ML, Loh WS. Surgical considerations for tracheostomy during the COVID-19 pandemic: lessons learned from the severe acute respiratory syndrome outbreak. JAMA Otolaryngol Head Neck Surg 2020 Jun 01;146(6):517-518. [doi: 10.1001/jamaoto.2020.0764] [Medline: 32232426]
12. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]

13.   China Industry Information. Chyxx.com. URL: http://www.chyxx.com/industry/202002/836124.html [accessed 2020-04-14]
14.   Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London 2020:1-20. [doi: 10.25561/77482]
15.   Abdullah AS, Tomlinson B, Cockram CS, Thomas GN. Lessons from the severe acute respiratory syndrome outbreak in Hong Kong. Emerg Infect Dis 2003 Sep;9(9):1042-1045 [FREE Full text] [doi: 10.3201/eid0909.030366] [Medline: 14519237]
16.   Coronavirus disease (COVID-19) advice for the public: when and how to use masks. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks [accessed 2020-02-14]
17.   Mokhtar F, Ramli D. Singapore u-turns on wearing masks as local virus cases climb. Bloomberg. 2020. URL: https://www.bloomberg.com/news/articles/2020-04-03/singapore-u-turns-on-wearing-masks-as-local-virus-cases-climb
18.   Recommendation regarding the use of cloth face coverings, especially in areas of significant community-based transmission. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover.html
19.   Javid B, Weekes MP, Matheson NJ. Covid-19: should the public wear face masks? BMJ 2020 Apr 09;369:m1442. [doi: 10.1136/bmj.m1442] [Medline: 32273278]
20.   Macintyre CR. Cluster randomised controlled trial: hand hygiene and face mask use within 36 hours of index patient symptom onset reduces flu transmission to household contacts. Evid Based Med 2010 Apr;15(2):48-49. [doi: 10.1136/ebm1043] [Medline: 20436121]
21.   Leung NHL, Chu D, Shiu E, Chan KH, McDevitt JJ, Hau BJP, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 2020 May;26(5):676-680. [doi: 10.1038/s41591-020-0843-2] [Medline: 32371934]
22.   van der Sande M, Teunis P, Sabel R. Professional and home-made face masks reduce exposure to respiratory infections among the general population. PLoS One 2008 Jul 09;3(7):e2618 [FREE Full text] [doi: 10.1371/journal.pone.0002618] [Medline: 18612429]
23.   Bradsher K, Alderman L. The world needs masks. China makes them, but has been hoarding them. The New York Times. 2020. URL: https://www.nytimes.com/2020/03/13/business/masks-china-coronavirus.html; [accessed 2020-04-04]

Abbreviations

AUROC: area under the receiver operating characteristics curve
NPI: nonpharmaceutical intervention
OR: odds ratio
RR: risk ratio

Original Paper

Associations Between COVID-19 Misinformation Exposure and Belief With COVID-19 Knowledge and Preventive Behaviors: Cross-Sectional Online Study

Jung Jae Lee[1], PhD; Kyung-Ah Kang[2], PhD; Man Ping Wang[1], PhD; Sheng Zhi Zhao[1], MPH; Janet Yuen Ha Wong[1], PhD; Siobhan O'Connor[3], PhD; Sook Ching Yang[4], MBChB; Sunhwa Shin[2], PhD

[1]School of Nursing, University of Hong Kong, Hong Kong, China (Hong Kong)

[2]College of Nursing, Sahmyook University, Seoul, Republic of Korea

[3]School of Health in Social Science, The University of Edinburgh, Edinburgh, United Kingdom

[4]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, United Kingdom

Corresponding Author:
Sunhwa Shin, PhD
College of Nursing
Sahmyook University
815 Hwarang-ro, Nowon-gu
Seoul, 01795
Republic of Korea
Phone: 82 2 3399 1595
Email: shinsh@syu.ac.kr

Abstract

Background: Online misinformation proliferation during the COVID-19 pandemic has become a major public health concern.

Objective: We aimed to assess the prevalence of COVID-19 misinformation exposure and beliefs, associated factors including psychological distress with misinformation exposure, and the associations between COVID-19 knowledge and number of preventive behaviors.

Methods: A cross-sectional online survey was conducted with 1049 South Korean adults in April 2020. Respondents were asked about receiving COVID-19 misinformation using 12 items identified by the World Health Organization. Logistic regression was used to compute adjusted odds ratios (aORs) for the association of receiving misinformation with sociodemographic characteristics, source of information, COVID-19 misinformation belief, and psychological distress, as well as the associations of COVID-19 misinformation belief with COVID-19 knowledge and the number of COVID-19 preventive behaviors among those who received the misinformation. All data were weighted according to the Korea census data in 2018.

Results: Overall, 67.78% (n=711) of respondents reported exposure to at least one COVID-19 misinformation item. Misinformation exposure was associated with younger age, higher education levels, and lower income. Sources of information associated with misinformation exposure were social networking services (aOR 1.67, 95% CI 1.20-2.32) and instant messaging (aOR 1.79, 1.27-2.51). Misinformation exposure was also associated with psychological distress including anxiety (aOR 1.80, 1.24-2.61), depressive (aOR 1.47, 1.09-2.00), and posttraumatic stress disorder symptoms (aOR 1.97, 1.42-2.73), as well as misinformation belief (aOR 7.33, 5.17-10.38). Misinformation belief was associated with poorer COVID-19 knowledge (high: aOR 0.62, 0.45-0.84) and fewer preventive behaviors (≥7 behaviors: aOR 0.54, 0.39-0.74).

Conclusions: COVID-19 misinformation exposure was associated with misinformation belief, while misinformation belief was associated with fewer preventive behaviors. Given the potential of misinformation to undermine global efforts in COVID-19 disease control, up-to-date public health strategies are required to counter the proliferation of misinformation.

Introduction

Background

COVID-19 has brought significant challenges to public health with its high infectivity and severity, particularly in vulnerable groups (eg, older adults, those with chronic diseases) [1-4], leading to a rapid rise in cases worldwide. On top of managing the response to the COVID-19 global health crisis, the World Health Organization (WHO) and governments also face the challenge of an "infodemic," which causes people to experience difficulties in finding credible and trustworthy sources amid an excess of information [5]. Although the concept of an infodemic is not a new one, the digital age we are currently living in has magnified its effects and added complexities to the challenge it poses. With the widespread use of social media such as internet websites, social networking services (SNSs), and instant messaging services, people all over the world are more connected than ever, allowing information to be shared easily and quickly [6]. However, a recent study has found that in the midst of the global pandemic, COVID-19 misinformation is just as likely to spread and engage users on social media platforms as accurate information, which can pose an equal threat to the COVID-19 public health response by affecting public awareness and knowledge of the disease [7].

Misinformation can be defined as information that is false or inaccurate and not supported by scientific evidence [8]. Misinformation in the context of COVID-19 can include inaccurate information regarding the virus and its transmission, conspiracy theories, and fabricated reports regarding methods of prevention and treatment [9]. Some of its consequences include the panic-buying and hoarding of goods, taking ineffective and potentially harmful remedies, ignoring advice from health authorities, and engaging in behavior that increases the risk of virus transmission [10]. Despite the many efforts by the WHO and public health organizations to battle the infodemic, such as conducting campaigns against COVID-19 misinformation, cooperating with social media platforms, and regularly providing evidence-based information to the public (eg, COVID-19 advice for the public: myth busters [11]), the proliferation of misinformation worldwide has remained rampant [9,12,13]. Although social media can be used effectively to provide essential health-related information to the global community, misinformation does not require professional verification or review, and thus has the potential to proliferate quicker and be disseminated farther on social media due to existing algorithms that highlight popular or desired content. This highlights the tall challenge health authorities face in delivering accurate information to the public in precedence to the proliferation of misinformation [4,5] and the need for new strategies to build preparedness [9] against future infodemics.

South Korea (hereafter Korea) has one of the most developed technological infrastructures in the world, and the use of the internet is well-integrated into Koreans' everyday lives. As such, information and communication technology (ICT) has also been employed in Korea's response to the COVID-19 pandemic, including the monitoring and tracking of COVID-19 cases, conveying of health information to professionals and the public, and for allocating and distributing resources, such as COVID-19 test kits and protective equipment [14,15]. With the deep integration of ICT such as social media into Koreans' daily lives and activities [6], it is expected that Koreans will be highly exposed to COVID-19 information and misinformation, through either active searches for information or passive receiving of information through messages, emails, or news feeds. Notwithstanding our awareness of misinformation and the risk it poses to public health, there remains little evidence on the prevalence of COVID-19 misinformation exposure and its effects on health beliefs and behaviors, including psychological well-being [16].

Objectives

We aimed to investigate the following: (1) the prevalence of COVID-19 misinformation exposure and misinformation belief, (2) associated factors including psychological distress with misinformation exposure; and (3) the associations between misinformation exposure and COVID-19 knowledge as well as preventive behaviors in Korean adults.

Methods

Study Design and Sampling

A cross-sectional online survey was conducted using the largest online survey platform in Korea [17]. This platform was chosen as it has 5 million survey panel members nationwide (as of 2020) and has been used to conduct more than 160,000 surveys for academic (such as those by Kim et al [18] and Ra et al [19]), government, and industry research. The inclusion criteria were the following: (1) aged ≥20 years (according to Korea's civil law, those aged ≥20 years are regarded as adults), (2) a resident in the Seoul Metropolitan area (including the Seoul, Gyeonggi-do, and Incheon areas, in which 50.0% of the Korean population resides as of 2020), (3) has encountered COVID-19 information from any source, and (4) a Korean speaker.

The company sent survey invitations containing general information about the survey such as its aim and participation incentive (KRW 1000 [US $1 is about KRW 1200]) via emails and a smartphone app to registered survey panel members who met the inclusion criteria on April 23, 2020. The survey closed on the same date (ie, recruitment was conducted for one day due to the cost involved). Details of the survey and consent statement were provided on the first page of the online survey. Respondents provided consent by clicking "Agree to participate in this survey" on the same page, before moving on to answer the survey questions. The survey took approximately 15 minutes to complete. The survey questionnaire is attached as Multimedia Appendix 1. Ethical approval was obtained from an institutional review board at Sahmyook University in Seoul (Ref: 2-1040781-A-N-012020021HR). Meanwhile, as of April 23, 2020, Korea had 10,708 confirmed cases of COVID-19 since the first case was reported on January 24, 2020. The daily new cases of COVID-19 peaked on March 3, 2020 (803 cases), after which there was a downward trend until April 23, 2020 (14 new cases).

Measurements

Sociodemographic characteristics including sex, age, education level, household arrangement, and monthly personal income were collected.

COVID-19 misinformation items used in this study were extracted from COVID-19 misinformation reports by the WHO [11,20], the main coordinator of the global COVID-19 pandemic response. In total, 12 misinformation items about COVID-19 transmission, infectivity, prevention, and treatment were included (see Question 2 in Multimedia Appendix 1). Respondents were allocated to one of two groups: misinformation exposure (defined as having seen one or more items of misinformation, through active searching or passive receiving means) or misinformation nonexposure. Respondents were then asked if they believed any of the 12 misinformation items to be correct (hereafter misinformation belief) or incorrect.

Measures for psychological distress included anxiety and depressive symptoms, using the Patient Health Questionnaire-4 (PHQ-4; four items), which consists of two subscales: the Generalized Anxiety Disorder-2 (GAD-2; two items) and Patient Health Questionnaire-2 (PHQ-2; two items) [21]. The score of each subscale ranges from 0 to 6, and a score of ≥3 indicates a high risk of anxiety (GAD-2) and depression (PHQ-2), respectively [21]. PHQ-4 was validated in Korean (Cronbach α=.79; acceptable convergent validity) [22]. An additional Primary Care Post-Traumatic Stress Disorder Screen for DSM-5 (PC-PTSD-5; five items) that was validated in Korean (Cronbach α=.87; acceptable concurrent validity) [23,24] was also adopted to screen respondents for posttraumatic stress disorder (PTSD) risk. The scores range from 0 to 5 and the cutoff score for high risk of PTSD is 3.

COVID-19 knowledge was assessed using five COVID-19 knowledge questions (definition, transmission modes, main symptoms, prevention, and treatment) that were extracted from a questionnaire developed by the WHO [25] (see Questions 4-8 in Multimedia Appendix 1). Higher scores indicate a higher COVID-19 knowledge level.

The number of COVID-19 preventive behaviors that the respondents performed during the past three months was assessed with 10 answer options that were extracted from the COVID-19 preventive methods recommended by the WHO [25] and Korea Centers for Disease Control and Prevention [26], such as washing hands regularly, covering one's mouth and nose when coughing, and social distancing (see Question 9 in Multimedia Appendix 1). Higher scores indicated a higher engagement in COVID-19 preventive behaviors.

Statistical Analysis

All data were weighted by age and sex distributions in the Seoul Metropolitan area, according to the Korea census data in 2018 [27]. Descriptive statistics were reported in numbers, proportions, means, and standard deviations (SD), as appropriate. The differences between the two groups (COVID-19 misinformation exposure group versus nonexposure group), including the respondents' sociodemographic characteristics, source of information, COVID-19 misinformation belief, psychological well-being (ie, anxiety, depressive, and PTSD symptoms), and COVID-19 knowledge and preventive behaviors were analyzed by chi-square test or t test, as appropriate. The responses on the knowledge and preventive behavior questions were categorized into binary groups according to the mean scores for the chi-square analysis. Logistic regression was used to compute odds ratios (ORs) and adjusted ORs (aORs) to identify the association of misinformation exposure (a binary variable) with the sociodemographic characteristics, source of information, and COVID-19 misinformation belief. The associations of COVID-19 misinformation belief with COVID-19 knowledge and COVID-19 preventive behaviors among the respondents who encountered misinformation were also investigated. As subgroup analyses, we included the interaction term to test if demographic characteristics modify the associations of COVID-19 knowledge, preventive behaviors, misinformation belief, and psychosocial distress following misinformation exposure. Sex, age, highest education level, household arrangement, and monthly personal income were adjusted for the adjusted regression models. STATA 15 (StataCorp LLC) was used to conduct all analyses.

Results

Of 1054 people who initially responded to the survey, five were excluded from the study as they reported that they had not encountered any COVID-19 information and therefore could not complete the survey. Among the 1049 respondents, 50.04% (n=505) were male, the mean age was 40.60 years (SD 12.87), 74.94% (n=786) had tertiary education or higher, 88.50% (n=929) lived with others, and 55.95% (n=587) had a monthly personal income <KRW 3,000,000 (US $2500; the average monthly income among those employed was KRW 2,970,000 [US $2475] in 2018; Table 1).

Table 1. Characteristics of COVID-19 misinformation exposure by respondents' demographics, and sociobehavioral and psychological symptoms (N=1049).

Variables	Participants, n (%)[a] N=1049	Weighted values[b], n (%)[a] N=1049	Misinformation Not exposed (n=338)	Exposed (n=711)	P value[c,d]
Sex					
Male	446 (42.52)	525 (50.04)	179 (53.08)	346 (48.60)	.15[c]
Female	603 (57.48)	524 (49.96)	158 (46.92)	366 (51.40)	.15[c]
Age (years), mean (SD)	39.51 (10.47)	43.60 (12.87)	45.86 (13.01)	42.53 (12.67)	.02[d]
20-29	198 (18.88)	196 (18.69)	48 (14.09)	149 (20.87)	.06[c]
30-39	351 (33.46)	219 (20.85)	65 (19.25)	154 (21.61)	.06[c]
40-49	311 (29.65)	243 (23.17)	78 (23.14)	165 (23.18)	.06[c]
50-59	148 (14.12)	236 (22.54)	74 (22.11)	162 (22.74)	.06[c]
60-69	41 (3.91)	154 (14.75)	72 (21.41)	82 (11.59)	.06[c]
Education					
High school or below	233 (22.21)	263 (25.06)	92 (27.13)	171 (24.08)	.83[c]
Tertiary or above	816 (77.79)	786 (74.94)	246 (72.87)	540 (75.92)	.83[c]
Household arrangement					
Living alone	142 (13.54)	121 (11.50)	32 (9.55)	88 (12.43)	.10[c]
Living with others	907 (86.46)	929 (88.50)	306 (90.45)	623 (87.57)	.10[c]
Monthly personal income, KRW (US $)					
<3,000,000 (<2500)[e]	622 (59.29)	587 (55.95)	165 (48.78)	422 (59.35)	.03[c]
3,000,000-4,990,000 (2500-4158)	264 (25.17)	261 (24.92)	99 (29.24)	163 (22.86)	.03[c]
≥5,000,000 (>4158)	163 (15.54)	201 (19.14)	74 (21.98)	127 (17.79)	.03[c]
COVID-19 information source[f]					
Television, radio, or newspaper (offline)	999 (95.23)	998 (95.07)	316 (93.53)	682 (95.80)	.40[c]
Television, radio, or newspaper (online)	1036 (98.76)	1030 (98.21)	328 (97.06)	703 (98.76)	.54[c]
Other internet websites	864 (82.36)	842 (80.26)	262 (77.55)	580 (81.55)	.44[c]
Social network services	753 (71.78)	754 (71.82)	220 (65.10)	534 (75.01)	.03[c]
Instant messaging	873 (83.22)	868 (82.71)	262 (77.49)	606 (85.19)	.011[c]
COVID-19 misinformation belief					
No	604 (57.58)	618 (58.86)	291 (86.21)	326 (45.87)	<.001[c]
Yes	445 (42.42)	432 (41.14)	47 (13.79)	385 (54.13)	<.001[c]
Anxiety symptom, mean (SD)[g]	1.49 (1.60)	1.51 (1.65)	3.18 (1.58)	3.66 (1.66)	<.001[d]
No	853 (81.32)	854 (81.38)	290 (85.94)	564 (79.22)	.002[c]
Yes	196 (18.68)	195 (18.62)	48 (14.06)	148 (20.78)	.002[c]
Depressive symptom, mean (SD)[h]	2.02 (1.73)	2.04 (1.78)	3.76 (1.78)	4.17 (1.77)	.001[d]
No	738 (70.35)	739 (70.45)	253 (74.85)	486 (68.36)	.01[c]
Yes	311 (29.65)	310 (29.55)	85 (25.15)	225 (31.64)	.01[c]

XSL•FO
RenderX

Variables	Participants, n (%)[a] N=1049	Weighted values[b], n (%)[a] N=1049	Misinformation Not exposed (n=338)	Exposed (n=711)	P value[c,d]
Posttraumatic stress disorder symptom, mean (SD)[i]	1.43 (1.09)	1.43 (1.14)	1.12 (1.04)	1.57 (1.15)	<.001[d]
No	902 (85.99)	903 (86.05)	306 (90.52)	597 (83.92)	.002[c]
Yes	147 (14.01)	146 (13.95)	32 (9.48)	114 (16.08)	.002[c]
COVID-19 knowledge, mean (SD)	24.69 (2.56)	24.65 (2.69)	24.79 (2.24)	24.58 (2.88)	.38[d]
Low (0-24)	444 (42.33)	446 (42.55)	142 (41.97)	305 (42.82)	.99[c]
High (25-35)	605 (57.67)	603 (57.45)	196 (58.03)	407 (57.18)	.99[c]
COVID-19 preventive behaviors, mean (SD)	6.94 (2.42)	7.01 (2.47)	7.01 (2.46)	7.02 (2.48)	.72[d]
0-6 behaviors	396 (37.75)	388 (36.99)	129 (38.15)	259 (36.44)	.70[c]
≥7 behaviors	653 (62.25)	661 (63.01)	209 (61.85)	452 (63.57)	.70[c]

[a]Calculated percentages were rounded off to one decimal place; accordingly, combined percentages can exceed 100%.

[b]Data were weighted by sex and age distribution of the general population in the Seoul metropolitan area.

[c]P for chi-square (computed using unweighted data).

[d]P for t test (computed using unweighted data).

[e]Average monthly income among employees was KRW 2,970,000 in 2018.

[f]Multiple responses allowed.

[g]Generalized Anxiety Disorder Questionnaire-2 (GAD-2) score ≥3.

[h]Patient Health Questionnaire-2 (PHQ-2) score ≥3.

[i]Primary Care Post-Traumatic Stress Disorder Screen for DSM-5 (PC-PTSD-5) score ≥3.

The majority of the respondents encountered COVID-19 information from diverse media including television, newspapers, internet websites, SNSs, and instant messaging. Overall, 57.45% (n=603) had high levels (score of 25-35) of COVID-19 knowledge and 63.01% (n=661) reported conducting ≥7 COVID-19 preventive behaviors. In total, 18.68% (n=196), 31.55% (n=331), and 13.95% (n=146) reported anxiety, depressive, and PTSD symptoms, respectively (Table 1).

Overall, 41.14% (n=432) reported believing in one or more of the 12 COVID-19 misinformation items, while exposure to at least one COVID-19 misinformation item in the preceding three months was reported by 67.78% (n=711) of the respondents. In addition, 49.76% (n=354) and 48.14% (n=342) encountered misinformation about reusing masks (Table 2), which was the most common item of misinformation reported.

XSL•FO
RenderX

Table 2. Respondents' exposure to COVID-19 misinformation (N=1049).

Variables	Participants, n (%)[a]	Weighted values[b], n (%)[a]
Number of COVID-19 misinformation items		
0	321 (30.60)	338 (32.19)
1	279 (26.60)	260 (24.74)
2	205 (19.54)	206 (19.63)
≥3 items	244 (23.26)	246 (23.44)
COVID-19 misinformation items[c]	728 (69.40)	711 (67.78)
Masks can be sterilized and reused after steaming with hot water	346 (47.53)	354 (49.76)
Masks can be reused after spraying alcohol on its surface	345 (47.39)	342 (48.14)
Drinking tea can prevent infection	231 (31.73)	219 (30.78)
Gargling can disinfect the respiratory tract to prevent infection	150 (20.60)	168 (23.55)
Coronavirus is artificially developed	172 (23.63)	156 (21.88)
Basking in the sun can prevent infection	110 (15.11)	109 (15.30)
Gargling with salt can prevent infection	104 (14.29)	109 (15.29)
Taking antibiotics can prevent or treat infection	94 (12.91)	102 (14.34)
Flip the sides of a used mask to reuse it	90 (12.36)	90 (12.60)
Drinking alcohol/smoking can prevent infection	59 (8.10)	70 (9.88)
Only older adults can be infected	25 (3.43)	32 (4.46)
A vaccine is available now	15 (2.06)	14 (1.99)

[a]Calculated percentages were rounded off to one decimal place; accordingly, combined percentages can exceed 100%.

[b]Data were weighted by sex and age distribution of the general population in the Seoul metropolitan area in Korea.

[c]The COVID-19 misinformation items were extracted from World Health Organization documents [11,20].

COVID-19 misinformation exposure was negatively associated with being older (60-69 years age group: aOR 0.40, 95% CI 0.25-0.64 versus 20-29 years age group) and having a higher monthly personal income (KRW 3,000,000-4,990,000 [US $2500-$4158]: aOR 0.66, 95% CI 0.47-0.93 versus <KRW 3,000,000 [US $2500]). COVID-19 misinformation exposure was positively associated with a tertiary education or above (aOR 1.42, 95% CI 1.02-1.96 versus high school or below). Of the information sources, misinformation exposure was associated with SNSs (aOR 1.75, 95% CI 1.31-2.35 versus other information sources) and instant messaging (aOR 1.79, 95% CI 1.27-2.51 versus other information sources), while offline and online television, radio, and newspapers and other websites were not statistically significant. Misinformation exposure was also significantly associated with misinformation belief (aOR 7.33, 95% CI 5.17-10.38 versus no misinformation belief) and with psychological distress, including anxiety (aOR 1.80, 95% CI 1.24-2.61 versus no anxiety symptoms), depressive (aOR 1.47, 95% CI 1.09-2.00 versus no depressive symptoms), and PTSD symptoms (aOR 1.97, 95% CI 1.42-2.73 versus no PTSD symptoms). However, misinformation exposure was not associated with COVID-19 knowledge and preventive behaviors (Table 3).

Table 3. Associated factors with COVID-19 misinformation exposure (N=1049)[a].

Variables	Misinformation exposure (yes/no)	
	Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)[b]
Sex		
Male	Reference	Reference
Female	1.20 (0.92-1.55)	1.15 (0.87-1.52)
Age (years)		
20-29	Reference	Reference
30-39	0.76 (0.49-1.17)	0.74 (0.46-1.17)
40-49	0.68 (0.44-1.03)	0.70 (0.45-1.10)
50-59	0.69 (0.45-1.06)	0.76 (0.48-1.20)
60-69	0.37 (0.23-0.58)[c]	0.40 (0.25-0.64)[c]
Education		
High school or below	Reference	Reference
Tertiary or above	1.17 (0.87-1.58)	1.42 (1.02-1.96)[d]
Household arrangement		
Living alone	Reference	Reference
Living with others (including family)	0.74 (0.49-1.14)	0.83 (0.53-1.29)
Monthly personal income, KRW (US $)[e]		
<3,000,000 (<2500)	Reference	Reference
3,000,000-4,990,000 (2500-4158)	0.64 (0.47-0.87)[f]	0.66 (0.47-0.93)[d]
≥5,000,000 (≥4158)	0.67 (0.47-0.93)[d]	0.74 (0.51-1.07)
COVID-19 information source[g]		
Television, radio, or newspaper (offline)	1.58 (0.89-2.78)	1.79 (0.99-3.22)
Television, radio, or newspaper (online)	2.41 (0.97-6.03)	2.52 (0.98-6.52)
Other internet websites	1.28 (0.93-1.76)	1.24 (0.89-1.72)
Social network services	1.61 (1.22-2.13)[f]	1.75 (1.31-2.35)[c]
Instant messaging	1.67 (1.20-2.32)[f]	1.79 (1.27-2.51)[f]
COVID-19 misinformation belief		
No	Reference	Reference
Yes	7.38 (5.24-10.39)[c]	7.33 (5.17-10.38)[c]
COVID-19 knowledge		
Low (0-24)	Reference	Reference
High (25-35)	0.97 (0.74-1.26)	0.97 (0.74-1.27)
COVID-19 preventive behaviors		
0-6 behaviors	Reference	Reference
≥7 behaviors	1.08 (0.82-1.41)	1.13 (0.86-1.49)
Anxiety symptom[h]		
No	Reference	Reference
Yes	1.60 (1.12-2.29)[f]	1.80 (1.24-2.61)[f]
Depressive symptom[i]		

Variables	Misinformation exposure (yes/no)	
	Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)[b]
No	Reference	Reference
Yes	1.38 (1.03-1.84)[d]	1.47 (1.09-2.00)[d]
Posttraumatic stress disorder symptom[j]		
No	Reference	Reference
Yes	1.84 (1.34-2.53)[c]	1.97 (1.42-2.73)[c]

[a]All data were weighted by sex and age distribution of the general population in the Seoul metropolitan area in Korea.

[b]Adjusted for sex, age, highest education level, household arrangement, and monthly personal income.

[c]$P<.001$.

[d]$P<.05$.

[e]Average monthly income among employees was KRW 2,970,000 in 2018.

[f]$P<.01$.

[g]Multiple responses were allowed (the reference groups were those who responded "No").

[h]Generalized Anxiety Disorder Questionnaire-2 (GAD-2) score ≥ 3.

[i]Patient Health Questionnaire-2 (PHQ-2) score ≥ 3.

[j]Primary Care Post-Traumatic Stress Disorder (PTSD) Screen for DSM-5 (PC-PTSD-5) score ≥ 3.

Subgroup analyses showed that the associations of misinformation exposure with misinformation belief, COVID-19 knowledge, preventive behaviors, and psychological distress differed according to the respondents' demographics (sex, age, education, and monthly personal income; Multimedia Appendix 2). Among the respondents who reported misinformation exposure, misinformation belief was associated with lower COVID-19 knowledge levels (high: aOR 0.62, 95% CI 0.45-0.84 versus low) and fewer COVID-19 preventive behaviors (≥ 7 behaviors: aOR 0.54, 95% CI 0.39-0.74 versus 0-6 behaviors; Table 4).

Table 4. Associations of COVID-19 knowledge and number of COVID-19 preventive behaviors with COVID-19 misinformation belief among respondents who were exposed to misinformation (N=711)[a].

Variables	Participants, n (%)	Misinformation belief (yes/no)	
		Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)[b]
Knowledge			
Low (0-24)	259 (36.44)	Reference	Reference
High (25-35)	452 (63.56)	0.59 (0.44-0.80)[c]	0.62 (0.45-0.84)[c]
Preventive behaviors			
0-6 behaviors	305 (42.82)	Reference	Reference
≥ 7 behaviors	407 (57.18)	0.51 (0.37-0.70)[d]	0.54 (0.39-0.74)[d]

[a]All data were weighted by sex and age distribution of the general population in the Seoul metropolitan area in Korea.

[b]Adjusted for sex, age, highest education level, household arrangement, and monthly personal income.

[c]$P<.01$.

[d]$P<.001$.

Discussion

Principal Findings

In this cross-sectional survey of Korean adults, more than two-thirds of the respondents reported COVID-19 misinformation exposure between the end of January 2020 and the end of April 2020, as COVID-19 evolved into a global pandemic. A previous study [28] identified a similar prevalence, where 70% of respondents reported misinformation exposure during the 2018 Ebola virus epidemic, affirming a substantial exposure risk to inaccurate or false health-related information during serious infectious disease outbreaks.

We identified that misinformation exposure was significantly associated with younger age, higher education levels, and lower incomes. Existing studies reported that younger people, including university students, preferred to obtain health information via online means and perceived themselves as having a high level of digital health literacy [29-31]. Such characteristics would expose young people to more COVID-19 misinformation and information. However, contrary to their perceptions, they lacked the skills to evaluate health resources

and apply gathered information to health-related decisions [29-31]. This indicates that despite their proficiencies in using technology and the internet, effective interventions are required to improve young people's digital health literacy, which is the ability to search for, understand, and critically evaluate health information through electronic sources, then apply gained knowledge to health issues [32]. Meanwhile, to our best knowledge, the associations between misinformation exposure and demographic characteristics have been underinvestigated in the existing literature. We performed subgroup analyses that offer additional details of interactions between the respondents' demographic characteristics and COVID-19 variables. Further studies that provide in-depth understanding on how and why these demographic characteristics are associated with misinformation exposure will be useful.

Consistent with previous reports on the role of social media in health information dissemination and misinformation propagation [7,9,33,34], respondents in this study reported greater COVID-19 misinformation exposure through SNSs and instant messaging. This can also be attributed to the significance of social media like SNSs and instant messaging in Koreans' daily lives, as those services also include product marketing, shopping, and payment services, contributing to Korea having the third highest social media usage penetration in the world [35].

Many recent studies have identified that the COVID-19 pandemic and fear of being infected have had negative effects on public mental health, reporting increased depression and anxiety [36-38]. The respondents in this study were similarly identified to be at high risk of psychological distress; compared to the nonexposure group, the misinformation exposure group notably had around 1.8 and 1.47 times higher anxiety and depressive symptoms, respectively. This demonstrates the alarming negative effect that misinformation can have on public mental health. Higher levels of social media use during the COVID-19 pandemic have been shown to result in higher levels of anxiety (OR 1.72) and a combination of depression and anxiety (OR 1.91) [39]. It is purported that prolonged and frequent use of social media throughout the ongoing pandemic increases exposure to misinformation along with accurate information. The mixture of accurate and false information can deliver conflicting messages and amplify uncertainties regarding COVID-19 and its perceived health risks [40], resulting in psychological distress [41,42]. A vicious cycle can be triggered, as evidence has shown that psychological distress itself can drive people to look for more information, which in turn causes further distress [43]. Despite the ongoing COVID-19 situation, this study also identified that the respondents had symptoms of PTSD in relation to the pandemic, raising concern that the psychological impact can persist and lead to poor physical health outcomes [44].

In this study, no association between misinformation exposure and COVID-19 knowledge as well as preventive behaviors was found. However, we identified that COVID-19 misinformation belief was negatively associated with COVID-19 knowledge and preventive behaviors, while it was positively associated with misinformation exposure. Similar to our findings, Allington et al [45] found that frequent use of SNSs, which propagate

more misinformation than any other media [7], for COVID-19 information was associated with having conspiracy beliefs. Conspiracy beliefs, in turn, showed a negative relationship with COVID-19 preventive behaviors [45-47]. Vinck et al [28] also reported that the belief in Ebola virus misinformation resulted in a lower likelihood of adopting preventive behaviors.

The Health Belief Model (HBM) is a theoretical framework widely used in public health to understand heath behaviors for disease prevention (see Champion and Skinner [48]). It theorizes that people's beliefs about their susceptibility to COVID-19 infection and its severity (collectively known as perceived threat), as well as perceptions about the benefits of and barriers to engaging in preventive behaviors will be predictive of their likelihood of engaging in those behaviors, while cues trigger engagement. Additionally, one's understanding about COVID-19 can alter individual beliefs and thus indirectly influence behavior [48]. Based on the HBM, our findings suggest that those who believed in misinformation that they were exposed to had lesser accurate knowledge of COVID-19, which could include inaccurate knowledge about preventive behaviors. A less accurate understanding of COVID-19 in turn altered the perceptions they had about COVID-19, such as reduced perceived COVID-19 threat, reduced perceived benefits from preventive behaviors, heightened perceived benefits from inappropriate preventive behaviors, or heightened perceived barriers to preventive behaviors. These altered perceptions and beliefs regarding COVID-19 thus resulted in reduced engagement in recommended preventive behavior, as our findings show, or potentially increased engagement in inappropriate preventive behaviors. The HBM thus offers insights into interventions that can improve engagement in recommended preventive behaviors by delivering information targeted at the core HBM beliefs of perceived susceptibility, severity, benefits, and barriers [49]. Although our findings show no association between misinformation exposure and COVID-19 knowledge, increasing exposure to accurate or corrective COVID-19 information can be useful to denormalize misinformation beliefs, thus changing perceptions of COVID-19 to increase the likelihood of engaging in recommended preventive behaviors [4].

Implications

As misinformation exposure is associated with misinformation belief, it is essential to manage and stem the propagation of misinformation through popular mediums like social media and counteract misinformation exposure with evidence-based information exposure. A study identified that official authorities had produced only a few COVID-19 information videos through a popular video streaming website (YouTube, a website with around 2 billion monthly users), while videos containing misinformation were disproportionately increasing [50]. Governments, health agencies, and researchers should take advantage of such social media outlets by producing and sharing evidence-based and corrective information through YouTube videos and simple but impactful infographics. Additionally, governments and health agencies can work closely with social media platforms to ensure that health-related information has increased visibility without involving unilateral censorship, develop misinformation alerts, and provide verification of

information source quality, particularly in the event of global health crises [51]. These can accordingly mitigate the development of misinformation belief.

Increasing digital health literacy among the public, particularly young people, will also be essential, as misinformation exposure did not reflect improved COVID-19 knowledge or preventive behaviors but was associated with misinformation belief. Educational or training programs on digital health literacy should be developed and delivered to the public and can be introduced in schools to cultivate these skills from a young age.

In-depth, follow-up, and longitudinal studies that explore misinformation selection and beliefs, as well as how misinformation beliefs transform health behaviors will be beneficial as foundations to digital health literacy program development and anti-misinformation strategies.

Limitations

This study has several limitations. First, the causal relationships between COVID-19 misinformation exposure and belief, COVID-19 knowledge, and psychological distress were uncertain due to the cross-sectional study design [52]. Second, we used survey questions to measure respondents' COVID-19 misinformation exposure, COVID-19 knowledge, and preventive behaviors, which are not validated. Third, we collected self-reported data from the respondents that would cause recall and social desirability biases. Fourth, although all data were weighted according to the South Korea census data, there were relatively fewer people aged ≥60 years who participated in this study. We recruited adult respondents only and did not recruit teenagers (ie, those aged <20 years), who popularly use social media for information acquisition. The inclusion of younger people in future studies would provide additional evidence about their misinformation exposure and belief. Fifth, we conducted an online survey by recruiting panel members using a survey company and thus there is a possibility of sampling bias. For instance, those sampled were based in urban, not rural, areas and drawn from a high-income Asian country with prior experience of managing outbreaks of infectious diseases (eg, Middle East Respiratory Syndrome).

Conclusion

We investigated the prevalence of misinformation exposure and factors that were associated with misinformation exposure and belief, including psychological distress, COVID-19 knowledge, and preventive behaviors in an adult population in the Seoul Metropolitan area, Korea. COVID-19 misinformation exposure was associated with misinformation belief, while misinformation belief was associated with poorer knowledge and engagement in fewer preventive behaviors. Given the potential of such misinformation to undermine global efforts in the COVID-19 response, public health strategies should be kept up-to-date and involve collaborations with multiple stakeholders, including social media platforms, to counter the proliferation of misinformation and win the fight against the infodemic.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 27 KB - jmir_v22i11e22205_app1.docx]

Multimedia Appendix 2
Interaction effect of demographic characteristics with COVID-19 knowledge, preventive behaviours, misinformation belief, and psychological distress on misinformation exposure.
[DOCX File , 32 KB - jmir_v22i11e22205_app2.docx]

References

1. COVID-19 dashboard 2020. Johns Hopkins University. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-06-12]
2. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):1 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
3. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
4. van der Meer TGLA, Jin Y. Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source. Health Commun 2020 May;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
5. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report 86. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf?sfvrsn=c615ea20_6 [accessed 2020-06-03]

6. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar 21;395(10228):942-943 [FREE Full text] [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]

7. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]

8. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]

9. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychol Sci 2020 Jul;31(7):770-780. [doi: 10.1177/0956797620939054] [Medline: 32603243]

10. Coronavirus: The psychology of panic buying. BBC. 2020. URL: https://www.bbc.com/worklife/article/20200304-coronavirus-covid-19-update-why-people-are-stockpiling [accessed 2020-05-01]

11. World Health Organization. Coronavirus disease (COVID-19) advice for the public: Myth busters. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-05-25]

12. World Health Organization. Countering misinformation about COVID-19. 2020. URL: https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19 [accessed 2020-06-06]

13. Tasnim S, Hossain MM, Mazumder H. Impact of Rumors and Misinformation on COVID-19 in Social Media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]

14. World Health Organization. Digital technology for COVID-19 response. 2020. URL: https://www.who.int/news-room/detail/03-04-2020-digital-technology-for-covid-19-response [accessed 2020-06-12]

15. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]

16. Editorial. Coronavirus misinformation needs researchers to respond. Nature 2020 May;581(7809):355-356. [doi: 10.1038/d41586-020-01550-y] [Medline: 32461657]

17. DOOIT Survey. URL: http://www.dooit.co.kr [accessed 2020-10-30]

18. Kim H, Min J, Kim H, Min K. Accident risk associated with smartphone addiction: A study on university students in Korea. J Behav Addict 2017 Dec 01;6(4):699-707 [FREE Full text] [doi: 10.1556/2006.6.2017.070] [Medline: 29099234]

19. Ra Y, Kim C, Kim Y, Han J. Survey of Anxiety in Ordinary Workers and Doctors Regarding Sedative Use during Endoscopic Examination in the Seoul Metropolitan Area. Gut Liver 2016 Sep 15;10(5):786-795 [FREE Full text] [doi: 10.5009/gnl16048] [Medline: 27563022]

20. World Health Organization. 2020. URL: https://www.who.int/westernpacific/news/q-a-detail/food-safety-and-nutrition [accessed 2020-08-20]

21. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]

22. Kim H, Shin C. Lee S-H, Han C: Standardization of the Korean version of the Patient Health Questionnaire-4 (PHQ-4). Clinical Psychopharmacology and Neuroscience 2020 Mar:1.

23. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. J Gen Intern Med 2016 Oct;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]

24. Jung Y, Kim D, Kim W, Roh D, Chae J, Park JE. A Brief Screening Tool for PTSD: Validation of the Korean Version of the Primary Care PTSD Screen for DSM-5 (K-PC-PTSD-5). J Korean Med Sci 2018 Dec 24;33(52):e338 [FREE Full text] [doi: 10.3346/jkms.2018.33.e338] [Medline: 30584416]

25. World Health Organization. Risk Communication and Community Engagement (RCCE) action plan guidance COVID-19: preparedness and response. 2020. URL: https://www.who.int/publications-detail/risk-communication-and-community-engagement-(rcce)-action-plan-guidance [accessed 2020-03-25]

26. Korea Centers for Disease Control and Prevention. COVID-19 prevention methods. 2020. URL: http://ncov.mohw.go.kr/baroView4.do?brdId=4&brdGubun=4 [accessed 2020-03-30]

27. National population census. Statistics Korea (KOSTAT). 2018. URL: http://kosis.kr/index/index.do [accessed 2020-05-04]

28. Vinck P, Pham PN, Bindu KK, Bedford J, Nilles EJ. Institutional trust and misinformation in the response to the 2018-19 Ebola outbreak in North Kivu, DR Congo: a population-based survey. Lancet Infect Dis 2019 May;19(5):529-536. [doi: 10.1016/S1473-3099(19)30063-5] [Medline: 30928435]

29. Robb M. Shellenbarger T: Influential factors and perceptions of eHealth literacy among undergraduate college students. Online Journal of Nursing Informatics 2014;18(3):1.

30. Kim S, Son Y. Relationships Between eHealth Literacy and Health Behaviors in Korean Adults. Comput Inform Nurs 2017 Feb;35(2):84-90. [doi: 10.1097/CIN.0000000000000255] [Medline: 27258808]

31. Heuberger RA, Ivanitskaya L. Preferred Sources of Nutrition Information: Contrasts Between Younger and Older Adults. Journal of Intergenerational Relationships 2011 Apr;9(2):176-190. [doi: 10.1080/15350770.2011.567919]

32. World Health Organization. Digital health literacy. 2017. URL: https://www.who.int/global-coordination-mechanism/working-groups/digital_hl.pdf [accessed 2020-08-29]

33. Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. Health Informatics J 2016 Sep;22(3):470-478. [doi: 10.1177/1460458214568037] [Medline: 25656678]

34. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]

35. Statista. Active social network penetration in selected countries as of January 2020. 2020. URL: https://www.statista.com/statistics/282846/regular-social-networking-usage-penetration-worldwide-by-country/ [accessed 2020-05-23]

36. Ornell F, Schuch JB, Sordi AO, Kessler FHP. "Pandemic fear" and COVID-19: mental health burden and strategies. Braz J Psychiatry 2020;42(3):232-235. [doi: 10.1590/1516-4446-2020-0008] [Medline: 32267343]

37. Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. Psychiatry Clin Neurosci 2020 Apr;74(4):281-282 [FREE Full text] [doi: 10.1111/pcn.12988] [Medline: 32034840]

38. Fitzpatrick KM, Drawve G, Harris C. Facing new fears during the COVID-19 pandemic: The State of America's mental health. J Anxiety Disord 2020 Oct;75:102291 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102291] [Medline: 32827869]

39. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

40. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]

41. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]

42. Rochwerg B, Parke R, Murthy S, Fernando SM, Leigh JP, Marshall J, et al. Misinformation During the Coronavirus Disease 2019 Outbreak: How Knowledge Emerges From Noise. Crit Care Explor 2020 Apr;2(4):e0098 [FREE Full text] [doi: 10.1097/CCE.0000000000000098] [Medline: 32426740]

43. Thompson RR, Jones NM, Holman EA, Silver RC. Media exposure to mass violence events can fuel a cycle of distress. Sci Adv 2019 Apr;5(4):eaav3502 [FREE Full text] [doi: 10.1126/sciadv.aav3502] [Medline: 31001584]

44. Garfin DR, Thompson RR, Holman EA. Acute stress and subsequent health outcomes: A systematic review. J Psychosom Res 2018 Sep;112:107-113. [doi: 10.1016/j.jpsychores.2018.05.017] [Medline: 30097129]

45. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7 [FREE Full text] [doi: 10.1017/S003329172000224X] [Medline: 32513320]

46. Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. Psychol Med 2020 May 21:1-13 [FREE Full text] [doi: 10.1017/S0033291720001890] [Medline: 32436485]

47. Allington D, Dhavan N. The relationship between conspiracy beliefs and compliance with public health guidance with regard to COVID-19. London: Centre for Countering Digital Hate URL: https://core.ac.uk/download/pdf/322486467.pdf [accessed 2020-10-30]

48. Champion V, Skinner CS. The Health Belief Model. In: Glanz K, Rimer B, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice, 4th edition. NJ, USA: John Wiley & Sons; 2008:45-65.

49. Romano JL. Politics of Prevention: Reflections From the COVID-19 Pandemic. Journal of Prevention and Health Promotion 2020 Jul 13;1(1):34-57. [doi: 10.1177/2632077020938360]

50. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May;5(5):1. [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]

51. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]

52. Carlson MDA, Morrison RS. Study design, precision, and validity in observational studies. J Palliat Med 2009 Jan;12(1):77-82 [FREE Full text] [doi: 10.1089/jpm.2008.9690] [Medline: 19284267]

Abbreviations

aOR: adjusted odds ratio
GAD-2: Generalized Anxiety Disorder-2
HBM: Health Belief Model
ICT: information and communication technology
OR: odds ratio
PHQ-4: Patient Health Questionnaire-4
PTSD: posttraumatic stress disorder

SNS: social networking service
WHO: World Health Organization

XSL•FO
RenderX

Original Paper

Psychological Impact of Health Risk Communication and Social Media on College Students During the COVID-19 Pandemic: Cross-Sectional Study

Mengyao Li[1], PhD; Li Liu[1], PhD; Yilong Yang[1], PhD; Yang Wang[1], PhD; Xiaoshi Yang[1], PhD; Hui Wu[1], PhD

China Medical University, Shenyang, China

Corresponding Author:
Hui Wu, PhD
China Medical University
No 77 Puhe Road
Shenyang North New Area
Shenyang, 110122
China
Phone: 86 18900910568
Email: hwu@cmu.edu.cn

Abstract

Background: The outbreak of COVID-19 began in 2019 and is expected to impact the psychological health of college students. Few studies have investigated the associations among health risk communication, social media, and psychological symptoms during a major pandemic.

Objective: The aim of this research was to assess the prevalence of psychological symptoms among college students and explore their associations with health risk communication and social media.

Methods: A web-based survey was distributed through the Wenjuanxing platform among Chinese college students from March 3-15, 2020. In addition to demographics, information on health risk communication and social media was collected, and the Symptom Checklist 90 Phobia and Health Anxiety Inventory subscale was used to assess psychological symptoms among 1676 college students in China. Multivariable logistic regression was performed to examine these independent risk factors.

Results: The prevalence of panic and health anxiety was 17.2% (288/1676) and 24.3% (408/1676), respectively. Regarding risk communication, understanding the risk of COVID-19 (odds ratio [OR] 0.480, 95% CI 0.367-0.627) was a protective factor against panic. Knowledge of prognosis (OR 0.708, 95% CI 0.551-0.910), preventive measures (OR 0.380, 95% CI 0.195-0.742), and wearing face masks (OR 0.445, 95% CI 0.230-0.862) were shown to be protective factors in predicting health anxiety. Perceived lethality (OR 1.860, 95% CI 1.408-2.459), being affected by the global spread (OR 1.936, 95% CI 1.405-2.669), and impact on social contacts (OR 1.420, 95% CI 1.118-1.802) were identified as significant risk factors associated with health anxiety. In terms of social media, trust in mainstream media (OR 0.613, 95% CI 0.461-0.816) was considered to be a protective factor against health anxiety.

Conclusions: There was a high prevalence of psychological symptoms among college students. Health risk communication and social media use were important in predicting psychological symptoms, especially health anxiety. Scientific and evidence-based information should be reported by social media platforms. Web-based consultation and intervention measures should be the focus of future studies.

Introduction

The COVID-19 pandemic is a global health threat [1]. The World Health Organization (WHO) has declared the COVID-19 outbreak to be a public health emergency of international concern [2]. The COVID-19 pandemic has spread worldwide and has had a significant impact on public health, governments, and social systems [3]. As of April 26, 2020, the number of global cumulative confirmed cases had exceeded 2,800,000, and the cumulative death toll was 192,971 [4].

XSL•FO
RenderX

The pandemic is not only leading to rapidly increasing numbers of confirmed cases and deaths but is also having a psychological impact on patients and health care workers who are exposed to COVID-19 [5]. Although most people are free from COVID-19 infection, they are being psychologically impacted by the COVID-19 pandemic [6,7]. People who are quarantined may have numerous psychological symptoms, such as anxiety, depression, panic, guilt, and stress [8]. Panic and anxiety are two common psychological symptoms experienced during the outbreak; this is likely related to the limited knowledge about COVID-19 and its highly infectious nature [8-11]. These psychological symptoms may cause adverse events and further contribute to social instability and racism [12].

College students are a key population of concern during the COVID-19 pandemic. The college years are a peak period for psychological symptom onset [13]. The high prevalence of COVID-19, home quarantine measures, shortage of masks during the initial stage of the pandemic, and delays in opening schools, colleges, and universities are expected to aggravate the psychological symptoms of college students. The mental health of college students is significantly affected during public health emergencies [14]. The WHO indicated that almost 90% of the world`s students (more than 1,500,000,000 children and young people) have been affected by nationwide school closures [15]. It is necessary to help college students maintain and develop their psychological health during the pandemic.

Researchers have increasingly acknowledged the importance of consistent, clear, and effective health risk communication throughout a pandemic [16]. Risk communication refers to an interactive process of exchanging information among individuals, groups, and institutions [17]. Health risk communication has a critical impact on the spread of pandemics and may help public health officials improve pandemic strategies and messaging [18]. Effective risk communication contributes to the implementation of precautionary behaviors, especially when new pandemic infectious disease risks arise in the phase when no treatment is available [19]. Based on protection motivation theory [20] and knowledge-attitude-practice theory [21], risk perception, correct knowledge, understanding, attitudes, and skills [16,22,23] are key to promoting and implementing preventive behavior [19,24]. Health risk communication is also important in maintaining psychological health. Emotional experiences during a hazardous event can influence individuals' evaluation of the negative outcomes of that event [25]. Previous studies have found that health risk communication can be a predictor of postdisaster mental health [26,27]. Risk perception is associated with mental symptoms, and dread of risk can increase the risk of mental symptoms [28]. Risk perception is associated with psychological health in earthquake survivors [29]. However, the association between health risk communication and psychological health during a pandemic has not yet been established, especially among college students.

Mass media plays a critical role during new and rapidly spreading global health risks [30]. In China, official departments provide daily updates about surveillance and active cases through social media [10]. This information helps the public improve their awareness of the outbreak. In addition, in a study by Gao et al [31], 82% of participants were found to frequently engage with social media. In addition to mainstream media, websites release and transfer COVID-19–related information [10]. The proliferation of internet-based health news encourages the referencing of media and academic articles, which may exaggerate the strength of results and mislead the public [32]. Contradictory, doubtful (untrustworthy), false, or misleading information may lead to public panic and, in turn, cause mental health symptoms [33]. These untrusted media sources heighten confusion and fearmongering [8,32] and cause anxiety and panic among students. As young people, college students are willing to obtain information on the internet. Therefore, college students may access more information sources and more complex content and thus may be more seriously affected [34,35]. This phenomenon may have a significant psychological impact on college students.

Thus, the COVID-19 pandemic may place a great psychological burden on college students worldwide. Health risk communication and trusted sources of information regarding COVID-19 are essential to prevent excessive panic and anxiety and to manage the outbreak in a scientific and effective way. The aims of this study are to explore the prevalence of panic and anxiety among college students and to identify the relationships among health risk communication, social media, and these two psychological symptoms. By identifying factors associated with negative psychological impacts, we hope that future research and interventions can be developed from our findings.

Methods

Ethics Statement

The procedures used in the current study were approved by the Committee on Human Experimentation of China Medical University (YDJK2020022). All subjects voluntarily enrolled in this research. Informed consent was provided by each participant.

Study Design

This cross-sectional survey was conducted from March 3-15, 2020. During this period, the total number of confirmed cases of COVID-19 exceeded 80,000 in China [36]. Chinese college students were invited to participate in a web-based survey though the Wenjuanxing platform. In total, 1676 college students participated in this study.

Measurement of Panic

Panic was measured using the Phobia subscale of the self-reported Symptom Checklist 90 (SCL-90) [37]. This scale consists of 7 items that mainly cover phobia symptoms, with an emphasis on situations with limited availability of help and avoidance behavior. The score for each item ranges from 0-4. If a score is >2 (above average), some phobia symptoms may be present [38]. The Cronbach α coefficient of the total scale was .925.

Measurement of Health Anxiety

The Health Anxiety Inventory (HAI) was used to measure health anxiety [39]. The HAI is an 18-item scale, and each item has 4

answer options ranging from 0 ("I do not") to 3 ("I spend most of my time") [39]. The presence of health anxiety is defined as a total score ≥15 [40]. The Cronbach α coefficient of the total scale was .818.

Measurement of Health Risk Communication of COVID-19

Health risk communication of COVID-19 was measured through two aspects: COVID-19–related perceptions and knowledge of preventive behaviors.

The COVID-19–related perceptions included knowledge of the prognosis of COVID-19, understanding of the risk of COVID-19, perceived lethality of COVID-19, perceived severity of COVID-19, feeling that it is difficult to protect oneself from being affected by the pandemic, and being affected by the global spread. Each item had 5 answer options: very low, low, moderate, high, and very high. The responses were subsequently categorized as moderate (moderate/low/very low) or high (high/very high).

Knowledge of preventive behaviors included knowledge of preventive measures, awareness of handwashing, awareness of wearing a face mask, and impact of home quarantine on social contacts. The questions regarding preventive measures had 5 response options: very low, low, moderate, high, and very high. The responses were subsequently categorized as moderate (moderate/low/very low) or high (high/very high). The questions about handwashing, wearing a face mask, and the impact of home quarantine on social contacts had 5 response options: strongly disagree, slightly disagree, moderate, slightly agree, and strongly agree. The responses were subsequently categorized as no (strongly disagree/slightly disagree/moderate) or yes (slightly agree/strongly agree).

Measurement of Social Media Sources of COVID-19 Information

Social media use was measured by asking which sources the participant used to obtain COVID-19–related information during the last month. These sources included WeChat, Sina Weibo, websites, television, newspapers, broadcast, government notices, or family members or friends. Trusted information sources were identified by asking "What do you think is the most trusted source of information regarding COVID-19?" The answers were divided into two groups: mainstream media and nonmainstream media.

Demographic Characteristics

The demographic factors examined included gender, major, monthly disposable income, region, number of cases in the participant's province, and year of study. Gender was divided into male and female. Major was divided into medicine-related major and other. Year was categorized as Year 1/2 and Year 3/4/5. Monthly disposable income was categorized into three groups: ≤1000 yuan (≤US $151), 1001-2000 yuan (US $162-303), and >2000 yuan (>US $303). Region was divided into rural and urban. Confirmed cases per province was divided into <1000 and ≥1000.

Statistical Analysis

First, comparisons between participants with and without panic or health anxiety were performed using chi-square tests for categorical variables. Next, a multivariable logistic regression analysis was conducted to identify which determinants contributed most to the likelihood of panic or health anxiety. Logistic regression analyses were used to explain the associations between the prevalence of panic, health anxiety, social media, and health risk communication after controlling for covariates. The aforementioned variables were all entered into the multivariate model. The Nagelkerke R^2 value was used as the coefficient of determination. Missing values were replaced with mean values. The Hosmer-Lemeshow test was used to examine the goodness-of-fit of the model, and a P value >.05 indicated acceptable fitness. Statistical analysis was conducted using SPSS 21.0 (IBM Corp), and a two-tailed P value <.05 was viewed as statistically significant.

Results

Participant Characteristics

Among the 1676 students, the prevalence of panic was 17.2% (n=288), and the prevalence of health anxiety was 24.3% (n=408). The mean age of the students was 20.17 years (SD 1.497). As shown in Tables 1-3, most of the college students were female (1088/1676, 64.9%). Most of the participants (1220/1676, 72.8%) were in Year 1 or Year 2. In total, 1235/1676 students (73.7%) were from urban areas. According to our results, only 121/1676 students (7.2%) were from Hubei, Guangdong, Henan, Zhejiang, or Hunan Province, each of which had more than 1000 confirmed cases of COVID-19.

Table 1. Demographic factors of the respondents and differences between panic and health anxiety (N=1676).

Demographic factor	Total, n (%)	Panic				Health anxiety			
		Yes, n (%)	No, n (%)	χ^2 [a]	P value	Yes, n (%)	No, n (%)	χ^2 [a]	P value
Gender				22.499	<.001			0.116	.73
Male	588 (35.1)	136 (8.1)	452 (27.0)			146 (8.7)	442 (26.4)		
Female	1088 (64.9)	152 (9.1)	936 (55.8)			262 (15.6)	826 (49.3)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Major				5.735	.02			0.123	.73
Medicine-related major	1219 (72.7)	193 (11.5)	1026 (61.2)			294 (17.5)	925 (55.2)		
Other	457 (27.3)	95 (5.7)	362 (21.6)			114 (6.8)	343 (20.5)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Year of study				4.539	.03			16.743	<.001
1/2	1220 (72.8)	195 (11.6)	1025 (61.2)			265 (15.8)	955 (57.0)		
3/4/5	456 (27.2)	93 (5.5)	363 (21.7)			143 (8.5)	313 (18.7)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Monthly disposable income (yuan)[b]				1.961	.38			4.479	.11
≤1000	366 (21.8)	67 (4.0)	299 (17.8)			103 (6.1)	263 (15.7)		
1001-2000	1038 (61.9)	182 (10.9)	856 (51.2)			236 (14.1)	802 (47.9)		
>2000	272 (16.2)	39 (2.3)	233 (13.9)			69 (4.1)	203 (12.1)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Region				3.229	.07			0.188	.66
Rural	441 (26.3)	88 (5.2)	353 (21.1)			104 (6.2)	337 (20.1)		
Urban	1235 (73.7)	200 (11.9)	1035 (61.8)			304 (18.1)	931 (55.5)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Confirmed cases in province				2.1	.15			4.406	.04
<1000	1555 (92.8)	273 (16.3)	1282 (76.5)			369 (22.0)	1186 (70.8)		
≥1000	121 (7.2)	1 (0.1)	106 (6.3)			39 (2.3)	82 (4.9)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		

[a]All degrees of freedom are 1 except for monthly disposable income, for which the degrees of freedom are 2.

[b]1 yuan=US $0.14 on March 3, 2020.

XSL•FO
RenderX

Table 2. Knowledge of COVID-19–related risks and protective measures reported by the respondents and differences between panic and health anxiety (N=1676).

Health risk knowledge	Total, n (%)	Panic				Health anxiety			
		Yes, n (%)	No, n (%)	χ^2_1	*P* value	Yes, n (%)	No, n (%)	χ^2_1	*P* value
Perception of COVID-19–related information									
Knowledge of prognosis				0.009	.92			15.479	<.001
Moderate	514 (30.7)	89 (5.3)	425 (25.4)			157 (9.4)	357 (21.3)		
High	1162 (69.3)	199 (11.9)	963 (57.5)			251 (15.0)	911 (54.4)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Understanding of the risk				36.618	<.001			3.601	.06
Moderate	782 (46.7)	181 (10.8)	601 (35.9)			207 (12.4)	575 (34.3)		
High	894 (53.3)	107 (6.3)	787 (47.0)			201 (12.0)	693 (41.3)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Perceived lethality				0.004	.95			49.315	<.001
Moderate	1300 (77.6)	223 (13.3)	1077 (64.3)			265 (15.8)	1035 (61.8)		
High	376 (22.4)	65 (3.9)	311 (18.6)			143 (8.5)	233 (13.9)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Feeling that it is difficult to protect oneself from the pandemic				4.658	.03			22.688	<.001
Moderate	1482 (88.4)	244 (14.6)	1238 (73.9)			334 (19.9)	1148 (68.5)		
High	194 (11.6)	44 (2.6)	150 (8.9)			74 (4.4)	120 (7.2)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Perceived severity				3.443	.06			21.551	<.001
Moderate	1228 (73.3)	225 (13.4)	1011 (60.3)			265 (15.8)	971 (57.9)		
High	448 (26.7)	63 (3.8)	377 (22.5)			143 (8.5)	297 (17.7)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Affected by global spread				2.579	.11			24.624	<.001
Moderate	399 (23.8)	58 (3.5)	341 (20.3)			60 (3.6)	339 (20.2)		
High	1277 (76.2)	230 (13.7)	1047 (62.5)			348 (20.8)	929 (55.4)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Knowledge of preventive behaviors									
Knowledge of preventive measures				3.566	.06			17.235	<.001
Moderate	43 (2.6)	12 (0.7)	31 (1.8)			22 (1.3)	21 (1.3)		
High	1633 (97.4)	276 (16.5)	1357 (81.0)			386 (23.0)	1247 (74.4)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Awareness of handwashing				10.416	.001			1.269	.26
No	70 (4.2)	22 (1.3)	48 (2.9)			21 (1.3)	49 (2.9)		
Yes	1606 (95.8)	266 (15.9)	1340 (80.0)			387 (23.1)	1219 (72.7)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Awareness of wearing face masks				4.452	.04			8.025	.005
No	45 (2.7)	13 (0.8)	32 (1.9)			19 (1.1)	26 (1.6)		
Yes	1631 (95.8)	275 (16.4)	1356 (80.9)			389 (23.2)	1242 (74.1)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Impact of home quarantine on social contacts				1.588	.21			18.277	<.001
No	1045 (62.4)	189 (11.8)	856 (51.1)			218 (13.0)	827 (49.3)		

Health risk knowledge	Total, n (%)	Panic				Health anxiety			
		Yes, n (%)	No, n (%)	χ^2_1	P value	Yes, n (%)	No, n (%)	χ^2_1	P value
Yes	631 (37.6)	99 (5.9)	532 (31.7)			190 (11.3)	441 (26.3)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		

Table 3. Number of social media information sources and trust in sources reported by the respondents and differences between panic and health anxiety (N=1676).

Social media source	Total, n (%)	Panic				Health anxiety			
		Yes, n (%)	No, n (%)	χ^2_1	P value	Yes, n (%)	No, n (%)	χ^2_1	P value
Number of information sources				1.082	.30			0.711	.40
≤3	745 (44.5)	136 (8.1)	609 (36.3)			174 (104)	571 (34.1)		
≥4	931 (55.5)	152 (9.0)	779 (46.5)			234 (14.0)	697 (41.6)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		
Trust in information				6.641	.01			20.757	<.001
Nonmainstream	301 (18)	67 (4.0)	234 (14.0)			104 (6.2)	197 (11.8)		
Mainstream	1375 (82)	221 (13.2)	1154 (67.9)			304 (18.1)	1071 (63.9)		
Total	1676 (100)	288 (17.2)	1388 (82.8)			408 (24.3)	1268 (75.7)		

Relationships Among Demographic Factors, Health Risk Communication, Social Media, and Panic

With respect to demographic factors, the groups showed significant differences in the distributions of gender, major, and year of study in the univariate analyses according to chi-square tests. No significant differences were found for monthly disposable income, region, or cases per province between the two groups.

Students who understood the risks of COVID-19 and felt less affected by the outbreak had lower panic levels. With respect to preventive knowledge, students who were aware that handwashing and wearing face masks can prevent COVID-19 had less panic. Regarding social media, students who trusted mainstream media information had less panic.

Relationships Among Demographic Factors, Social Media, Health Risk Communication, and Health Anxiety

With respect to demographic factors, the groups showed significant differences in the distributions of year of study and cases per province in the univariate analyses according to chi-square tests. No significant differences were found for gender, major, monthly disposable income, or region between the two groups.

Regarding COVID-19–related perception, students who had knowledge of the prognosis of the disease and felt less affected by the outbreak had lower health anxiety. Students who thought COVID-19 was lethal and severe and who were affected by the global spread had more health anxiety. In terms of preventive knowledge, students who had knowledge of preventive measures and were aware that wearing face masks could prevent COVID-19 had less health anxiety. Students who considered that their social contacts were affected by home quarantine had more health anxiety. Regarding social media, participants who believed mainstream information was more trustworthy had less health anxiety.

Risk Factors for Panic

Multivariable logistic regression analysis was conducted to identify which determinants contributed most to the likelihood of developing panic. The results of the unadjusted model and a model adjusting for potentially confounding demographic factors are reported. The confounders included in the adjusted model were major, gender, and year. The Nagelkerke pseudo-R^2 values for the unadjusted and adjusted models were 0.048 and 0.074, respectively. The Hosmer-Lemeshow tests demonstrated adequate fitness for the unadjusted (χ^2_4=2.733, P=.60) and adjusted (χ^2_8=5.790, P=.67) models. The variables in the adjusted model explained 7.4% of the variance in panic. The main effects of the adjusted model were similar to the crude results except for the awareness of handwashing. As shown in the adjusted model, a better understanding of the risk of COVID-19 (odds ratio [OR] 0.480, 95% CI 0.367-0.627) was identified as a significant protective factor for panic. These data are shown in Table 4.

XSL•FO
RenderX

Table 4. Results of logistic regression of risk factors for panic.

Variable	Unadjusted model[a]		Adjusted model[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
COVID-19–related perception				
Understanding the risk of COVID-19 (high vs moderate)	0.476 (0.366-0.621)	<.001	0.480 (0.367-0.627)	<.001
Felt it was difficult to protect oneself from the pandemic (high vs moderate)	1.279 (0.881-1.855)	.20	1.243 (0.853-1.812)	.26
Knowledge of preventive behaviors				
Awareness of handwashing (yes vs no)	0.533 (0.292-0.975)	.04	0.619 (0.336-1.143)	.13
Awareness of wearing face masks (yes vs no)	0.892 (0.412-1.942)	.77	0.933 (0.427-2.034)	.86
Information source				
Trusted information (mainstream media vs nonmainstream media)	0.738 (0.538-1.013)	.06	0.792 (0.574-1.092)	.15
Covariates				
Gender (female vs male)	N/A[d]	N/A	0.526 (0.403-0.687)	<.001
Year of study (3/4/5 vs 1/2)	N/A	N/A	1.268 (0.951-1.692)	.11
Major (medicine-related major vs other)	N/A	N/A	0.720 (0.538-0.962)	.03

[a]Nagelkerke R^2=0.048.

[b]Nagelkerke R^2=0.074.

[c]OR: odds ratio.

[d]N/A: not applicable.

Risk Factors for Health Anxiety

The results of the unadjusted model and a model adjusting for potentially confounding demographic factors are reported in Table 5. The confounders included in the adjusted model were cases per province and year of study. The Nagelkerke pseudo-R^2 values for the unadjusted and adjusted models were 0.110 and 0.119, respectively. The Hosmer-Lemeshow tests demonstrated adequate fitness for the unadjusted (χ^2_8=3.610, *P*=.89) and adjusted (χ^2_7= 3.080, *P*=.88) models. The variables in the adjusted model explained 11.9% of the variance in health anxiety. The main effects of the adjusted model were similar to the crude results except for perceived severity. As shown in the adjusted model, knowledge of prognosis (OR 0.708, 95% CI 0.551-0.910), knowledge of preventive measures (OR 0.380, 95% CI 0.195-0.742), awareness of wearing a face mask (OR 0.445, 95% CI 0.230-0.862), and trust in mainstream media (OR 0.613; 95% CI 0.461-0.816) were shown to be protective factors in predicting health anxiety. Perceived lethality (OR 1.860, 95% CI 1.408-2.459), being affected by the global spread (OR 1.936, 95% CI 1.405-2.669), and impact on social contacts (OR 1.420, 95% CI 1.118-1.802) were significant risk factors for health anxiety.

Table 5. Results of logistic regression of risk factors for health anxiety.

Variable	Unadjusted model[a]		Adjusted model[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
COVID-19–related perception				
Knowledge of prognosis (high vs moderate)	0.698 (0.544-0.896)	.005	0.708 (0.551,0.910)	.007
Perceived lethality (high vs moderate)	1.898 (1.438-2.506)	<.001	1.860 (1.408,2.459)	<.001
Felt it was difficult to protect oneself from the pandemic (high vs moderate)	1.231 (0.866-1.750)	.25	1.232 (0.865-1.753)	.25
Perceived severity (high vs moderate)	1.309 (1.003-1.708)	.047	1.302 (0.996-1.701)	.05
Affected by global spread (high vs moderate)	2.014 (1.464-2.772)	<.001	1.936 (1.405-2.669)	<.001
Knowledge of preventive behaviors				
Knowledge of preventive measures (high vs moderate)	0.374 (0.193-0.728)	.004	0.380 (0.195-0.742)	.005
Awareness of wearing face masks (yes vs no)	0.439 (0.228-0.844)	.014	0.445 (0.230-0.862)	.02
Impact of home quarantine on social contacts (yes vs no)	1.434 (1.131-1.818)	.003	1.420 (1.118-1.802)	.004
Social media				
Trusted information (mainstream media vs nonmainstream media)	0.611 (0.460-0.812)	0.001	0.613 (0.461-0.816)	.001
Covariates				
Number of confirmed cases in province (≥1000 vs <1000)	N/A[d]	N/A	1.405 (0.927-2.219)	.11
Year of study (3/4/5 vs 1/2)	N/A	N/A	1.453 (1.129-1.870)	.004

[a]Nagelkerke R^2=0.110.

[b]Nagelkerke R^2=0.119.

[c]OR: odds ratio.

[d]N/A: not applicable.

Discussion

Principal Findings

This study investigated the prevalence of panic and health anxiety among college students and explored the associations of health risk communication and social media with panic and health anxiety during the pandemic outbreak. Our results indicated that the prevalence of panic was 17.2% (288/1676) and the prevalence of health anxiety was 24.3% (408/1676). Previous studies have also reported that pandemics can trigger psychological symptoms [31,41,42]. If a pandemic constitutes an uncertain and threatening situation, it is more likely to trigger psychological symptoms. During pandemics, the number of people whose mental health is affected is greater than the number of people infected with the disease [43]. It is necessary to implement psychological interventions for college students during the COVID-19 pandemic.

Health risk communication was found to be important in predicting psychological symptoms among college students. With respect to risk perception, understanding the risk of COVID-19 was the only influencing factor for panic. Understanding the risk of COVID-19 could help relieve panic among college students. Similarly, knowledge of the prognosis

of COVID-19 was a protective factor in predicting health anxiety. Individuals behave in a more reticent and conservative manner when they feel threatened by disease [44]. Awareness of risk may help students take effective measures to prevent infection and avoid panic and anxiety. Understanding the risk, prognosis, and routes of COVID-19 infection further decreases panic and health anxiety. Our results are in line with previous studies. Receiving more health information is correlated with lower levels of psychological distress [45]. Properly understanding information is important for reducing negative psychological responses brought on by inaccurate perceptions [46]. Clear communication involving regular, accurate updates on the COVID-19 outbreak plays a critical role in developing psychological health [6]. At present, there are still limited effective treatments and vaccines for COVID-19, and the high infectivity, lethality, and global spread of the disease are causing health anxiety among college students.

Knowledge of preventive behaviors was another aspect of health risk communication that significantly predicted health anxiety among college students. Similarly, knowledge of preventive measures and of wearing a face mask were protective factors in predicting health anxiety. Accurate knowledge helped individuals react to and positively combat the outbreak, and it resulted in less negative emotion. Knowledge and guidance

about preventive behaviors are important factors in mitigating the spread of COVID-19 [47] and allaying unrealistic or excessive psychological anxiety [48]. Almost 37.6% of students (631/1676) felt that their social contacts were impacted by the home quarantine measures. The loss of freedom and increase of boredom had marked effects. Limitation of social contacts has a series of negative effects on psychological health. The etiology of anxiety as an illness includes a number of interacting biological, psychological, and social factors [49]. Individuals with better social networks are less likely to report anxiety symptoms [50]. Using class-based social groups may be able to improve college students' positive psychology and promote a positive atmosphere to enhance strong-tie relationships [51]. Even during the pandemic outbreak, social support plays a critical role in alleviating students' negative psychological symptoms, including anxiety [14]. Students can keep in touch with their friends or relatives on the web or by telephone to maintain social connections.

Social media played a critical role in psychological health. Students who believed that mainstream information was more trustworthy experienced less anxiety. Social media was the primary means of distributing information. Based on our results, 55% of students obtained COVID-19–related information in more than 4 ways. Mass media exposure to "infomedia" through social media platforms can create anxiety because rumors, "fake news," and conspiracy theories make it difficult to find trustworthy information [52]. Misinformation has caused anxiety and even hampered the response to the outbreak. Students spend a lot of time on the internet, and they are more likely to be misguided and experience triggering of anxiety symptoms. Appropriate guidance from authorities, meanwhile, can prevent individuals from overreacting to the disease and engaging in excessively avoidant behaviors [53]. Official public health organizations provide accurate information on measures to avoid COVID-19 [54], and the information is considered trustworthy and reliable [55]. During the pandemic, most people want to receive information from municipal health services, health care providers, and official media sources [56]. In the face of public health emergencies, accurate and authoritative information is important for relieving psychological symptoms among college students.

Implications

There are several implications of this study for clinical practitioners and policy makers. First, more attention should be paid to the psychological health of college students during the COVID-19 pandemic, and protective measures must be increased. Regardless of whether they were infected, students experienced psychological impacts from the outbreak. The delayed college start, uncertainty, and potential negative impact on academic progression may enhance the psychological burden on college students. Second, social media should be held responsible for providing correct and evidence-based information. Our study indicated that 55.5% (931/1676) of students obtained COVID-19–related information from more than four sources. Social media reporting can have both positive and negative consequences, and it had a strong influence on the psychological health of students. Information on social media platforms should be managed to quickly reduce the spread of fear and uncertainty and enhance public trust in public health measures [52]. Third, social isolation should be avoided. Imposed quarantine, including separation from friends or relatives and a departure from usual daily routines, is an unpleasant experience [57]. Additionally, social isolation caused an increase in anxiety. The need for social support has increased during the current pandemic. It is necessary to communicate socially via the internet or by telephone during the COVID-19 pandemic. Finally, health risk communication is essential during the outbreak. Based on our research, almost all students possessed knowledge related to the prevention of COVID-19. While the severity of the pandemic can trigger psychological symptoms in college students, risk perception promotes appropriate practices among students [58]. Our results indicated that 95.8% of college students (1606/1676) believed that preventive behaviors were effective. These behaviors provided the students with a sense of security and decreased their fear and anxiety.

Limitations

Several limitations should be considered in this study. First, conclusions on causality cannot be drawn due to the cross-sectional design. Second, given the use of a web-based survey, there may be some response bias. Third, the explained variance was low.

Conclusion

Psychological symptoms among college students were found to be at high levels during the COVID-19 pandemic. Understanding the risk of COVID-19 was a protective factor for panic. Trust in mainstream media, knowledge of preventive measures of COVID-19, and knowledge of its prognosis were protective factors for anxiety. However, perceived lethality, the global spread, and impact on social contacts were risk factors for health anxiety. Effective health risk communication and scientific and evidenced-based information should be reported through social media. The psychological health of college students should be considered. Future research should focus on intervention measures to ensure college students' psychological well-being during a global pandemic outbreak.

Authors' Contributions

MYL and HW conceived and designed the investigation. LL, YW, and XSY performed the investigation. MYL and YLY analyzed the data. MYL wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
2. Mahase E. China coronavirus: WHO declares international emergency as death toll exceeds 200. BMJ 2020 Jan 31;368:m408. [doi: 10.1136/bmj.m408] [Medline: 32005727]
3. Pan X, Ojcius DM, Gao T, Li Z, Pan C, Pan C. Lessons learned from the 2019-nCoV epidemic on prevention of future infectious diseases. Microbes Infect 2020 Mar;22(2):86-91 [FREE Full text] [doi: 10.1016/j.micinf.2020.02.004] [Medline: 32088333]
4. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-04-26]
5. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
6. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiat 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/S2215-0366(20)30046-8] [Medline: 32032543]
7. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiat 2020 Apr;7(4):300-302 [FREE Full text] [doi: 10.1016/S2215-0366(20)30073-0] [Medline: 32085840]
8. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
9. Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. Psychiatry Clin Neurosci 2020 Apr;74(4):281-282 [FREE Full text] [doi: 10.1111/pcn.12988] [Medline: 32034840]
10. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
11. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
12. Devakumar D, Shannon G, Bhopal SS, Abubakar I. Racism and discrimination in COVID-19 responses. Lancet 2020 Apr 11;395(10231):1194 [FREE Full text] [doi: 10.1016/S0140-6736(20)30792-3] [Medline: 32246915]
13. de Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. Epidemiol Psychiatr Sci 2012 Mar;21(1):47-57. [doi: 10.1017/s2045796011000746] [Medline: 22670412]
14. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
15. Coronavirus disease 2019 (COVID-19) Situation Report-77. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200406-sitrep-77-covid-19.pdf?sfvrsn=21d1e632_2 [accessed 2020-04-06]
16. Jehn M, Kim Y, Bradley B, Lant T. Community knowledge, risk perception, and preparedness for the 2009 influenza A/H1N1 pandemic. J Public Health Manag Pract 2011;17(5):431-438. [doi: 10.1097/PHH.0b013e3182113921] [Medline: 21788781]
17. National Research Council (US) Committee on Risk Perception and Communication. Improving risk communication. Washington, DC: National Academy Press; 1989.
18. Henrich N, Holmes B. The public's acceptance of novel vaccines during a pandemic: a focus group study and its application to influenza H1N1. Emerg Health Threats J 2009;2:e8 [FREE Full text] [doi: 10.3134/ehtj.09.008] [Medline: 22460289]
19. Brug J, Aro AR, Richardus JH. Risk perceptions and behaviour: towards pandemic control of emerging infectious diseases: international research on risk perception in the control of emerging infectious diseases. Int J Behav Med 2009;16(1):3-6 [FREE Full text] [doi: 10.1007/s12529-008-9000-x] [Medline: 19127440]
20. Rogers RW. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In: Cacioppo JT, Petty R, editors. Social Psychophysiology: A Sourcebook. New York, NY: Guilford Press; 1983:153-177.
21. Ajilore K, Atakiti I, Onyenankeya K. College students' knowledge, attitudes and adherence to public service announcements on Ebola in Nigeria: Suggestions for improving future Ebola prevention education programmes. Health Educ J 2017 Jun 15;76(6):648-660. [doi: 10.1177/0017896917710969]

22. Murakami M, Nakatani J, Oki T. Evaluation of Risk Perception and Risk-Comparison Information Regarding Dietary Radionuclides after the 2011 Fukushima Nuclear Power Plant Accident. PLoS One 2016;11(11):e0165594 [FREE Full text] [doi: 10.1371/journal.pone.0165594] [Medline: 27802304]

23. Hano MC, Baghdikian CL, Prince S, Lazzarino E, Hubbell B, Sams E, et al. Illuminating Stakeholder Perspectives at the Intersection of Air Quality Health Risk Communication and Cardiac Rehabilitation. Int J Environ Res Public Health 2019 Sep 26;16(19):3603. [doi: 10.3390/ijerph16193603] [Medline: 31561473]

24. Renner B, Schupp H. The perception of health risks. In: The Oxford Handbook of Health Psychology. New York, NY: Oxford University Press; 2011.

25. Siegrist M, Sütterlin B. Human and nature-caused hazards: the affect heuristic causes biased decisions. Risk Anal 2014 Aug;34(8):1482-1494. [doi: 10.1111/risa.12179] [Medline: 24576178]

26. Tseng MM, Lin Y, Hu F, Cheng T. Risks perception of electromagnetic fields in Taiwan: the influence of psychopathology and the degree of sensitivity to electromagnetic fields. Risk Anal 2013 Nov;33(11):2002-2012. [doi: 10.1111/risa.12041] [Medline: 23551091]

27. McArdle SC, Rosoff H, John RS. The dynamics of evolving beliefs, concerns emotions, and behavioral avoidance following 9/11: a longitudinal analysis of representative archival samples. Risk Anal 2012 Apr;32(4):744-761. [doi: 10.1111/j.1539-6924.2012.01814.x] [Medline: 22500650]

28. Nukui H, Murakami M, Midorikawa S, Suenaga M, Rokkaku Y, Yabe H, et al. Mental Health and Related Factors of Hospital Nurses. Asia Pac J Public Health 2017 Mar;29(2_suppl):161S-170S. [doi: 10.1177/1010539516682589] [Medline: 28330404]

29. Xu J, Dai J, Rao R, Xie H. The association between exposure and psychological health in earthquake survivors from the Longmen Shan Fault area: the mediating effect of risk perception. BMC Public Health 2016 May 18;16:417 [FREE Full text] [doi: 10.1186/s12889-016-2999-8] [Medline: 27193051]

30. Schmälzle R, Häcker F, Renner B, Honey CJ, Schupp HT. Neural correlates of risk perception during real-life risk communication. J Neurosci 2013 Jun 19;33(25):10340-10347 [FREE Full text] [doi: 10.1523/JNEUROSCI.5323-12.2013] [Medline: 23785147]

31. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

32. Ippolito G, Hui DS, Ntoumi F, Maeurer M, Zumla A. Toning down the 2019-nCoV media hype-and restoring hope. Lancet Respir Med 2020 Mar;8(3):230-231 [FREE Full text] [doi: 10.1016/S2213-2600(20)30070-9] [Medline: 32146924]

33. Ashrafi-Rizi H, Kazempour Z. Information Typology in Coronavirus (COVID-19) Crisis; a Commentary. Arch Acad Emerg Med 2020;8(1):e19 [FREE Full text] [Medline: 32185370]

34. Cheng C, Jun H, Liang B. Psychological health diathesis assessment system: a nationwide survey of resilient trait scale for Chinese adults. Stud Psychol Behav 2014;12:735-742.

35. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]

36. The latest situation of the novel coronavirus pneumonia epidemic as of 24:00 on March 15. Webpage in Chinese. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/xcs/yqtb/202003/114113d25c1d47aabe68381e836f06a8.shtml [accessed 2020-11-13]

37. Derogatis L. SCL-90-R: Administration, Scoring, and Procedures Manual. Minneapolis, MN: NCS Pearson; 1996.

38. Wei Y, Li H, Wang H, Zhang S, Sun Y. Psychological Status of Volunteers in a Phase I Clinical Trial Assessed by Symptom Checklist 90 (SCL-90) and Eysenck Personality Questionnaire (EPQ). Med Sci Monit 2018 Jul 17;24:4968-4973 [FREE Full text] [doi: 10.12659/MSM.909524] [Medline: 30015333]

39. Salkovskis PM, Rimes KA, Warwick HMC, Clark DM. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002 Jul;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]

40. Zhang Y, Liu R, Li G, Mao S, Yuan Y. The reliability and validity of a Chinese-version Short Health Anxiety Inventory: an investigation of university students. Neuropsychiatr Dis Treat 2015;11:1739-1747 [FREE Full text] [doi: 10.2147/NDT.S83501] [Medline: 26213472]

41. Liao Q, Cowling BJ, Lam WWT, Ng DMW, Fielding R. Anxiety, worry and cognitive risk estimate in relation to protective behaviors during the 2009 influenza A/H1N1 pandemic in Hong Kong: ten cross-sectional surveys. BMC Infect Dis 2014 Mar 27;14:169 [FREE Full text] [doi: 10.1186/1471-2334-14-169] [Medline: 24674239]

42. Ji D, Ji Y, Duan X, Li W, Sun Z, Song X, et al. Prevalence of psychological symptoms among Ebola survivors and healthcare workers during the 2014-2015 Ebola outbreak in Sierra Leone: a cross-sectional study. Oncotarget 2017 Feb 21;8(8):12784-12791 [FREE Full text] [doi: 10.18632/oncotarget.14498] [Medline: 28061463]

43. Reardon S. Ebola's mental-health wounds linger in Africa. Nature 2015 Mar 05;519(7541):13-14. [doi: 10.1038/519013a] [Medline: 25739606]

44. Schaller M, Park J, Kenrick D. Human evolution and social cognition. In: Oxford Handbook of Evolutionary Psychology. Oxford, UK: Oxford University Press; 2007.

45. Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]

46. Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]

47. Basch CH, Hillyer GC, Meleo-Erwin ZC, Jaime C, Mohlman J, Basch CE. Preventive Behaviors Conveyed on YouTube to Mitigate Transmission of COVID-19: Cross-Sectional Study. JMIR Public Health Surveill 2020 Apr 02;6(2):e18807 [FREE Full text] [doi: 10.2196/18807] [Medline: 32240096]

48. Levin PJ, Gebbie EN, Qureshi K. Can the health-care system meet the challenge of pandemic flu? Planning, ethical, and workforce considerations. Public Health Rep 2007;122(5):573-578 [FREE Full text] [doi: 10.1177/003335490712200503] [Medline: 17877303]

49. McCutcheon VV. Toward an Integration of Social and Biological Research. Soc Serv Rev 2006 Mar;80(1):159-178 [FREE Full text] [doi: 10.1086/499087] [Medline: 20414476]

50. Levula A, Harré M, Wilson A. The Association Between Social Network Factors with Depression and Anxiety at Different Life Stages. Community Ment Health J 2018 Aug;54(6):842-854. [doi: 10.1007/s10597-017-0195-7] [Medline: 29127564]

51. Chang S, Lin Y, Lin C, Chang H, Chong PP. Promoting positive psychology using social networking sites: a study of new college entrants on Facebook. Int J Environ Res Public Health 2014 Apr 29;11(5):4652-4663 [FREE Full text] [doi: 10.3390/ijerph110504652] [Medline: 24785540]

52. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]

53. Schaller M, Murray DR, Bangerter A. Implications of the behavioural immune system for social behaviour and human health in the modern world. Philos Trans R Soc Lond B Biol Sci 2015 May 26;370(1669):20140105 [FREE Full text] [doi: 10.1098/rstb.2014.0105] [Medline: 25870392]

54. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]

55. Covolo L, Mascaretti S, Caruana A, Orizio G, Caimi L, Gelatti U. How has the flu virus infected the Web? 2010 influenza and vaccine information available on the Internet. BMC Public Health 2013 Jan 29;13:83 [FREE Full text] [doi: 10.1186/1471-2458-13-83] [Medline: 23360311]

56. van der Weerd W, Timmermans DR, Beaujean DJ, Oudhoff J, van Steenbergen JE. Monitoring the level of government trust, risk perception and intention of the general public to adopt protective measures during the influenza A (H1N1) pandemic in The Netherlands. BMC Public Health 2011 Jul 19;11:575 [FREE Full text] [doi: 10.1186/1471-2458-11-575] [Medline: 21771296]

57. Usher K, Bhullar N, Jackson D. Life in the pandemic: Social isolation and mental health. J Clin Nurs 2020 Aug;29(15-16):2756-2757. [doi: 10.1111/jocn.15290] [Medline: 32250493]

58. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]

Abbreviations

HAI: Health Anxiety Inventory
OR: odds ratio
SCL-90: Symptom Checklist 90
WHO: World Health Organization

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Threat, Coping, and Social Distance Adherence During COVID-19: Cross-Continental Comparison Using an Online Cross-Sectional Survey

Abrar Al-Hasan[1], BSc, MBA, PhD; Jiban Khuntia[2], ME, PhD; Dobin Yim[3], BS, MS, MBA, MA, PhD

[1]Kuwait University, AlShadadiya University City, Kuwait

[2]University of Colorado Denver, Denver, CO, United States

[3]Loyola University Maryland, Baltimore, MD, United States

Corresponding Author:
Abrar Al-Hasan, BSc, MBA, PhD
Kuwait University
College of Business Administration
AlShadadiya University City, 13055
Kuwait
Phone: 965 51165005
Email: abrar.alhasan@ku.edu.kw

Abstract

Background: Social distancing is an effective preventative policy for COVID-19 that is enforced by governments worldwide. However, significant variations are observed in adherence to social distancing across individuals and countries. Due to the lack of treatment, rapid spread, and prevalence of COVID-19, panic and fear associated with the disease causes great stress. Subsequent effects will be a variation around the coping and mitigation strategies for different individuals following different paths to manage the situation.

Objective: This study aims to explore how threat and coping appraisal processes work as mechanisms between information and citizens' adherence to COVID-19–related recommendations (ie, how the information sources and social media influence threat and coping appraisal processes with COVID-19 and how the threat and coping appraisal processes influence adherence to policy guidelines). In addition, this study aims to explore how citizens in three different countries (the United States, Kuwait, and South Korea), randomly sampled, are effectively using the mechanisms.

Methods: Randomly sampled online survey data collected by a global firm in May 2020 from 162 citizens of the United States, 185 of Kuwait, and 71 of South Korea were analyzed, resulting in a total sample size of 418. A seemingly unrelated regression model, controlling for several counterfactuals, was used for analysis. The study's focal estimated effects were compared across the three countries using the weighted distance between the parameter estimates.

Results: The seemingly unrelated regression model estimation results suggested that, overall, the intensity of information source use for the COVID-19 pandemic positively influenced the threat appraisal for the disease ($P<.001$). Furthermore, the intensity of social media use for the COVID-19 pandemic positively influenced the coping appraisal for the disease ($P<.001$). Higher COVID-19 threat appraisal had a positive effect on social distancing adherence ($P<.001$). Higher COVID-19 coping appraisal had a positive effect on social distancing adherence ($P<.001$). Higher intensity of COVID-19 knowledge positively influenced social distancing adherence ($P<.001$). There were country-level variations. Broadly, we found that the United States had better results than South Korea and Kuwait in leveraging the information to threat and coping appraisal to the adherence process, indicating that individuals in countries like the United States and South Korea may be more pragmatic to appraise the situation before making any decisions.

Conclusions: This study's findings suggest that the mediation of threat and coping strategies are essential, in varying effects, to shape the information and social media strategies for adherence outcomes. Accordingly, coordinating public service announcements along with information source outlets such as mainstream media (eg, TV and newspaper) as well as social media (eg, Facebook and Twitter) to inform citizens and, at the same time, deliver balanced messages about the threat and coping appraisal is critical in implementing a staggered social distancing and sheltering strategy.

XSL•FO
RenderX

KEYWORDS

COVID-19; adherence; coping appraisal; threat appraisal; protection motivation theory; social distancing; information sources; social media; knowledge; coping; threat; protection; motivation; cross-sectional; survey

Introduction

COVID-19 has instilled fear among all individuals across the world. Since beginning in Wuhan, China in December 2019, the COVID-19 pandemic has led to more than 6.5 million cases and 390,000 deaths reported worldwide as of June 4, 2020 [1].

Although there is no treatment or vaccine for COVID-19 [2] yet, the mitigation approach to COVID-19 has focused on infection control, effective quarantine, and treatment cure rates [3,4]. Infected individuals do not often exhibit any symptoms, and the disease progresses swiftly and kills patients at a much higher rate than the typical flu [5]. Limited testing availability, combined with few treatment options other than a ventilator to assist breathing, forces individuals to adopt a varying level of preventive measures [6-9], including measures to help alleviate psychological impacts [10,11].

Individuals can practice good respiratory hygiene by washing hands with soap and water often for at least 20 seconds and avoiding touching the eyes, nose, or mouth with unwashed hands. More drastic efforts include social distancing or maintaining a safe distance from others and sheltering practices by staying at home to avoid all contacts [2].

There are variations in citizens' adherence to the COVID-19–related social distancing guidelines suggested by governments across countries to manage and mitigate the disease, aligning to the preventive measures [12], for example, whether or not to wear a mask [13,14]. Citizen's willingness to follow the guidance will depend on the fear and anxiety caused by the disease. Arguably, compared to other diseases, infectious diseases like COVID-19 induce fear because of known and unknown reasons associated with spread and prevalence, the high rates of morbidity and mortality, and subsequent societal stigma and discrimination associated with the diagnosis and treatment of the disease [15-18].

Along with the actual nature and impact of COVID-19, the information available through different channels and the discussions through social media have played a significant role in influencing people's mindsets. The COVID-19 pandemic is associated with sleep disturbance and suicidal thoughts [19], misinformation and distress [20,21], and greater self-confidence when receiving information from more sources but only for health care workers [22]. Thus, the culmination of information and social media influences can influence the fear and coping strategies related to the disease and subsequent adherence to the policy recommendations [23].

The panic and fear associated with the disease will create a struggle in people's mindsets, including creating worst-case scenarios around their situations [24]. Subsequent effects will be a variation around the coping and mitigation strategies for different individuals following different paths to manage the situation.

In this context, this study asks two research questions: (1) How do the information sources and social media influence the threat and coping appraisal process with COVID-19? and (2) How does the threat and coping appraisal process influences adherence to policy guidelines?

The COVID-19 situation poses a sense of helplessness in that there is no treatment, and thus, individuals rely on different information sources to learn about the disease. Because they cannot see a doctor as a preassessment or plan for the disease, the information gained from the press, television, and internet influences their disease management and mitigation strategy. Thus, individual behavior would be affected by these information sources, as individuals will try to understand and maneuver complex situations [25-28]. Furthermore, given that social media is emerging as a critical information source to influence individual beliefs and perceptions, it is beyond doubt that such an influence would be quite useful in the COVID-19 situation [29-31].

Threat and coping appraisal mechanisms, grounded in the protection motivation theory (PMT), have been suggested to be precursors of individuals' actions to protect against infectious diseases [32] such as the influenza outbreak [33]. The intention to adopt a protective behavior such as adhering to the social distance recommendations results from perceiving a given threat such as a disease and desiring to avoid the adverse outcomes of such a threat [34]. The protection motivation perspective suggests that health risk is appraised by considering what the threat is because of the severity of a disease or health issue (threat appraisal: severity), how vulnerable an individual perceives the disease or health issue (threat appraisal: vulnerability), how successful preventative behavior is (coping appraisal: response efficacy), and how confident the individual feels in preventing the risk (coping appraisal: self-efficacy). Although several extensions and applications of the protection motivation concept are applied to health contexts, the basic premise of threat appraisal, consisting of severity and vulnerability, and coping appraisal, consisting of self and response efficacies, have remained similar (for a meta-analysis, see [34,35]).

Taking a closer look at people's actions, thoughts, or emotions can be a complicated process. However, resisting or suppressing our emotions creates paradoxical outcomes such as aggravating our fears instead of making them vanish. Thus, these coping strategies embrace and approach our stressors head-on to build grit and resilience. It is normal to find it challenging to be consistent when starting a new routine. In this context, exploration of COVID-19 mitigation using threat and coping appraisal mechanisms is essential.

Based on these discussions, this study focuses explicitly on testing the following hypotheses:

1. The intensity of the use of information sources for COVID-19 positively influences the threat appraisal for the disease.
2. The intensity of social media use for COVID-19 positively influences the coping appraisal for the disease.
3. A higher COVID-19 threat appraisal positively influences citizens' adherence intentions.
4. A higher COVID-19 coping appraisal positively influences citizens' adherence intentions.

This study conducts a comparative evaluation of the citizens' adherence process to COVID-19–related recommendations by the governments in three different countries: the United States, Kuwait, and South Korea.

Methods

Recruitment

This study started with a discussion in a focus group in Kuwait. The ten people who participated in the focus group opined that assessing threat and coping strategies is essential to manage the COVID-19 situation. Along with this insight, the group also suggested that different cultural systems and relevant mindsets will differ in adhering to government recommendations. This motivated us to study the research question across different countries with polarized mindsets and different cultural systems. Thus, the study expanded to other countries with different cultures. Due to resource constraints, the sampling was limited to countries that the authors have firsthand experience in explaining the similarities and differences.

A global survey-deploying firm collected the data for this study using online platforms. The firm recruited respondents from the United States, Kuwait, and South Korea in May 2020. The firm sampled respondents using an age, gender, ethnicity, and geographic region–based strata and quota matching process. Participation in the survey was free and voluntary; the respondents filled in electronic informed consent that was shown on the first page of the survey. The firm protects the confidentiality of anonymous respondents.

Data Collection

Data was collected using a survey instrument, as shown in Multimedia Appendix 1 Table A1. The questions asked participants about the cause and current state of the COVID-19 situation, their opinion on the government's role during the COVID-19 pandemic, the use of health information sources and social media for COVID-19–related information, and PMT measures for the COVID-19 pandemic adapted from previously validated scales [36-39]. The survey items included simple information-seeking questions and several existing validated scales from prior studies [35,40-47].

The survey instrument was pilot-tested using a sample of 48 respondents, leading to minor refinements to a few items. A total of 482 participants took the survey. Because of missing responses to the items, 64 observations were excluded, resulting in a sample size of 418. Responses were coded, validated, and analyzed using Stata version 16 (StataCorp).

Sample Demographics

Table 1 shows the descriptive statistics and pairwise correlations among the key variables used in this study. Out of 418 participants, 299 (58.7%) were female. The sample's largest age group was 18-27 years (n=192, 37.3%). This group was followed by the 28-37 years age group (n=150, 29.1%), 38-47 years age group (n=70, 13.6%), 48-57 years age group (n=43, 8.4%), and 58 years or older group (n=60, 11.7%). Multimedia Appendix 1 Figure A1 shows the countrywise comparison of the respondents' age. In terms of income level, 102 (20.5%) participants make less than US $30,000 annually; 92 (18.5%) make US $30,000-US $50,000; 102 (20.5%) make US $50,000-US $80,000; 65 (13.1%) make US $80,000-US $100,000; 57 (11.5%) make US $100,000-US $150,000; and 79 (15.9%) make more than US $150,000. The household income distribution as varied by country is shown in Multimedia Appendix 1 Figure A2. A detailed distribution of several demographic controls used in the models is available in Multimedia Appendix 1 Table A2.

Table 1. Summary statistics and pairwise correlations among key variables (N=418).

Variabless	Mean (SD)	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13
Adherence (1)	4.18 (1.00)	1	5	1.00												
Threat appraisal (2)	3.32 (0.87)	1	5	0.38	1.00											
Severity (3)	3.55 (0.97)	1	5	0.45	0.87	1.00										
Vulnerability (4)	3.10 (1.01)	1	5	0.21	0.89	0.55	1.00									
Coping appraisal (5)	2.93 (0.78)	1	5	0.53	0.44	0.46	0.32	1.00								
Self-efficacy (6)	3.06 (1.51)	1	5	0.28	0.31	0.30	0.25	0.84	1.00							
Response efficacy (7)	4.29 (0.89)	1	5	0.64	0.42	0.46	0.27	0.77	0.30	1.00						
Knowledge (8)	15.04 (3.62)	4	24	0.33	0.10	0.18	0.01	0.33	0.17	0.39	1.00					
Information source (9)	2.26 (1.25)	0	5	0.16	0.08	0.08	0.05	0.17	0.04	0.27	0.19	1.00				
Social media (10)	1.65 (1.43)	0	7	0.03	0.17	0.17	0.12	0.08	0.13	0.00	0.02	0.40	1.00			
Age (11)	2.27 (1.35)	1	5	0.00	0.00	0.03	–0.03	0.00	–0.10	0.12	0.09	0.21	0.04	1.00		
Female (12)	0.59 (0.49)	0	1	0.11	0.11	0.09	0.11	0.15	0.17	0.07	0.13	0.09	0.08	0.00	1.00	
Household income (13)	2.24 (1.72)	0	5	0.05	–0.07	–0.03	–0.09	0.13	0.14	0.06	0.11	0.04	–0.04	0.04	–0.12	1.00

Study Variables

The main dependent variable in this study is *adherence*. As shown in Multimedia Appendix 1 Table A1, *adherence* was measured using three questions of whether they would comply with the social distancing measures. The items' internal consistencies were tested using Cronbach alpha (0.81), and the standardized score was generated for the *adherence* variable. Table 1 shows that, on average, *adherence* is 4.18 out of 5, showing that most people are adhering to social distancing recommendations. The mean adherence level in Kuwait was the highest (mean 4.53, SD 0.81), followed by the United States (mean 4.14, SD 0.92) and South Korea (mean 3.53, SD 1.18). Multimedia Appendix 1 Figure A3 displays the mean adherence across countries.

Three main independent variables were of interest in this study to examine *adherence*. First, the independent variable *threat appraisal* consists of both the severity and vulnerability of the situation [46]. *Severity* is the perceived degree of harm from engaging in unhealthy behavior, the extent to which one will experience or die from contracting COVID-19 upon not following social distancing recommendations. *Vulnerability* is the perceived probability of threat occurrence, the extent to which one will contract COVID-19 upon not following social distancing recommendations. Both variables were operationalized using three questions adopted from previous studies to assess the *severity* and *vulnerability* of the COVID-19 pandemic (see Multimedia Appendix 1 Table A1). The items' internal consistency was tested using Cronbach alpha (0.63 and 0.76, respectively), and the standardized score was generated for both *severity* and *vulnerability* variables.

The second main independent variable was *coping appraisal*, which consists of both self-efficacy and response efficacy of the situation [46]. *Self-efficacy* is the perceived belief that one can successfully maintain a safe distance from others when in contact or stay home to avoid all contacts. *Response efficacy* is the perceived efficacy of adherence that adopting social

distancing will be effective in reducing the threat of COVID-19. Both variables were operationalized using three questions adopted from previous studies to assess the *self-efficacy* and *response efficacy* of adherence to the COVID-19 pandemic regulations (see Multimedia Appendix 1 Table A1). The items' internal consistency was tested using Cronbach alpha (0.6 and 0.8, respectively), and the standardized score was generated for both *self-efficacy* and *response efficacy* variables.

Lastly, the *knowledge* variable was coded to reflect the respondents' overall knowledge of COVID-19, as displayed in Multimedia Appendix 1 Table A1. The mean for *knowledge* was 15.04, with a minimum of 4 and a maximum of 24. The mean COVID-19 *knowledge* was highest in Kuwait (mean 16.32, SD 3.05), followed by the United States (mean 15.04, SD 3.62) and South Korea (mean 11.85, SD 3.74). Multimedia Appendix 1 Figure A4 displays the mean knowledge across countries.

To examine *threat appraisal* and *coping appraisal*, this study focuses on *COVID-19 information sources* used and on *COVID-19 social media* use. *COVID-19 information sources* was operationalized as the total number of sources used to attain COVID-19 health information. Table 1 displays the mean for the whole sample as 2.26, showing that individuals use two information sources on average to attain COVID-19 information. The mean was highest in Kuwait (mean 2.50, SD 1.13), followed by the United States (mean 2.31, SD 1.34), and South Korea (mean 1.68, SD 1.11). *COVID-19 social media* was also operationalized as the total number of social media platforms used to attain COVID-19 health information. Table 1 displays the mean for the whole sample as 1.16, showing that individuals use one social media platform to attain COVID-19 information. The mean was highest in Kuwait (mean 2.10, SD 1.38), followed by South Korea (mean 1.48, SD 1.36) and the United States (mean 1.29, SD 1.41). *COVID-19 social media* was further examined by categorizing the platforms to social network platforms (Facebook and LinkedIn), media sharing platforms (Instagram, Snapchat, TikTok, and YouTube), and texting and

microblogging platforms (Twitter and Whatsapp). Details of the analysis are shown in Multimedia Appendix 1 Table A5.

In addition to these key variables of interest, several control variables such as age, gender, and household income were included to account for counterfactual explanations relevant to our models (see Multimedia Appendix 1 Table A1 for details).

Econometric Analysis

Following PMT, the empirical model specifies how individuals express their opinion toward *adherence* of social distancing guidance by the government through *threat appraisal*, *coping appraisal*, and *knowledge* of COVID-19. Furthermore, *threat appraisal* and *coping appraisal* were specified through COVID-19 sources of information (*COVID-19 information sources* and *COVID-19 social media*). A set of control variables to enhance our empirical model's robustness included demographics characteristics of the survey participants, such as *gender*, *age*, and *household income*. The formal specification of the general model is as follows:

Threat appraisal model:

$$\textit{Threat appraisal}_i = \beta_0 + \beta_1 \times \textit{COVID-19 information sources}_i + \beta_2 \times \textit{COVID-19 social media}_i + \beta_3 \times \textit{Country} + \textit{Controls}_i + i \textbf{ (1)}$$

Coping appraisal model:

$$\textit{Coping appraisal}_i = \beta_0 + \beta_1 \times \textit{COVID-19 information sources}_i + \beta_2 \times \textit{COVID-19 social media}_i + \beta_3 \times \textit{Country} + \textit{Controls}_i + i \textbf{ (2)}$$

Adherence model:

$$\textit{Adherence}_i = \beta_0 + \beta_1 \times \textit{Threat appraisal\_hat}_i + \beta_2 \times \textit{Coping appraisal\_hat}_i + \beta_3 \times \textit{Knowledge}_i + \beta_4 \times \textit{Country} + \textit{Controls}_i + i \textbf{ (3)}$$

Where *controls* includes *gender*, *age groups*, and *household income*. The *country* dummy was included in the full sample model but was removed for subsample analyses.

Seemingly unrelated regression (SUR) was used to estimate the β coefficients of the key parameters and employ robust standard errors to test the models. are the disturbances associated with each observation. SUR was used to estimate to what extent our set of key variables influence *adherence*. The adherence model (Equation 3) uses the predicted values of threat appraisal (*threat appraisal_hat*) and coping appraisal (*coping appraisal_hat*) from the first stage models (Equations 1 and 2).

Results

Table 2 presents the key estimation results for Equations 1-3 for the whole sample. The first column (1) in the table shows the parameter estimates for the *coping appraisal* dependent variable, column (2) shows the parameter estimates for the *threat appraisal* dependent variable, and column (3) shows the parameter estimates for the *adherence* dependent variable for the full sample. Table 3 displays the key estimation results for Equations 1 and 2 for the individual countries. Columns 1-3 show the parameter estimates for the *coping appraisal* dependent variable, and columns 4-6 display the parameter estimates for the *threat appraisal* dependent variable for the United States, South Korea, and Kuwait, respectively. Table 4 presents the key estimation results for the *adherence* dependent variable, as in Equation 3, for the United States, South Korea, and Kuwait. Multimedia Appendix 1 Table A3 further analyzes the *adherence* model using the constituent variables (*severity*, *vulnerability*, *self-efficacy*, and *response efficacy*).

Table 2. Seemingly unrelated regression model results for the full sample.

Variables	DV[a]: coping appraisal (1)		DV: threat appraisal (2)		DV: adherence (3)	
	Full sample	P value	Full sample	P value	Full sample	P value
COVID-19 information source	0.112 (0.03)[b]	<.001	0.034 (0.04)	.35	N/A[c]	N/A
COVID-19 social media	0.034 (0.03)	.35	0.112 (0.03)	<.001	N/A	N/A
Threat appraisal	N/A	N/A	N/A	N/A	0.215 (0.05)	<.001
Coping appraisal	N/A	N/A	N/A	N/A	0.511 (0.06)	<.001
Knowledge	N/A	N/A	N/A	N/A	0.061 (0.01)	<.001
Age	−0.008 (0.03)	.72	0.011 (0.03)	.72	0.006 (0.03)	.85
Female	0.213 (0.08)	.04	0.174 (0.09)	.04	0.038 (0.08)	.64
Household income	0.055 (0.02)	.15	−0.037 (0.03)	.15	−0.002 (0.02)	.94
Constant	2.328 (0.13)	<.001	2.936 (0.15)	<.001	1.034 (0.23)	<.001
Observations, n	418	N/A	418	N/A	418	N/A
R^2	0.076	N/A	0.051	N/A	0.375	N/A
Chi-square	34.25	N/A	22.37	N/A	249.90	N/A
Root mean square error	0.7482	N/A	0.8420	N/A	N/A	N/A
P value	N/A	<.001	N/A	<.001	N/A	<.001

[a]DV: dependent variable.

[b]Standard errors in parentheses.

[c]N/A: not applicable.

Table 3. Coping and threat appraisal seemingly unrelated regression model results for individual countries.

Variables	DV[a]: coping appraisal						DV: threat appraisal					
	US (1)	P value	South Korea (2)	P value	Kuwait (3)	P value	US (4)	P value	South Korea (5)	P value	Kuwait (6)	P value
COVID-19 information source	0.099 (0.04)[b]	.01	0.221 (0.08)	.007	0.053 (0.04)	N/A[c]	0.021 (0.05)	.68	0.213 (0.12)	.07	0.077 (0.05)	.11
COVID-19 social media	−0.007 (0.04)	.86	−0.029 (0.08)	.71	−0.009 (0.03)	N/A	0.173 (0.05)	<.001	−0.122 (0.11)	.28	0.011 (0.04)	.77
Age	−0.0908 (0.04)	.01	−0.002 (0.08)	.98	0.045 (0.03)	N/A	−0.043 (0.05)	.35	0.040 (0.11)	.72	0.052 (0.04)	.19
Female	0.032 (0.11)	.76	0.058 (0.18)	.75	0.095 (0.08)	N/A	0.065 (0.14)	.63	−0.122 (0.26)	.63	0.246 (0.10)	.02
Household income	−0.030 (0.03)	.34	0.015 (0.07)	.82	0.029 (0.02)	N/A	−0.0917 (0.04)	.02	0.036 (0.10)	.71	−0.055 (0.03)	.06
Constant	2.652 (0.18)	<.001	1.726 (0.29)	<.001	3.042 (0.14)	N/A	2.907 (0.23)	<.001	2.883 (0.42)	<.001	3.212 (0.18)	<.001
Observations, n	162	N/A	71	N/A	185	N/A	162	N/A	71	N/A	185	N/A
R^2	0.075	N/A	0.117	N/A	0.048	N/A	0.139	N/A	0.050	N/A	0.063	N/A
Chi-square	13.20	N/A	9.47	N/A	9.27	N/A	26.18	N/A	3.77	N/A	12.43	N/A
Root mean square error	0.6307	N/A	0.6834	N/A	0.4996	N/A	0.8166	N/A	0.9817	N/A	0.6581	N/A
P value	N/A	.02	N/A	.09	N/A	.10	N/A	<.001	N/A	.58	N/A	.03

[a]DV: dependent variable.

[b]Standard errors in parentheses.

[c]N/A: not applicable.

XSL•FO

RenderX

Table 4. Adherence seemingly unrelated regression model results for individual countries.

Variables	DV[a]: adherence					
	US (1)	P value	South Korea (2)	P value	Kuwait (3)	P value
Threat appraisal	0.223 (0.08)[b]	.006	0.576 (0.12)	<.001	0.0504 (0.08)	.55
Coping appraisal	0.334 (0.11)	.002	0.531 (0.16)	.001	0.536 (0.11)	<.001
Knowledge	0.053 (0.02)	.01	0.047 (0.03)	.07	0.040 (0.02)	.03
Age	0.038 (0.05)	.42	–0.012 (0.08)	.88	–0.038 (0.04)	.39
Female	–0.025 (0.14)	.86	0.202 (0.17)	.24	0.043 (0.12)	.72
Household income	0.013 (0.04)	.77	–0.044 (0.06)	.49	–0.011 (0.03)	.74
Constant	1.644 (0.45)	<.001	–0.135 (0.39)	.73	1.968 (0.52)	<.001
Observations, n	162	N/A[c]	71	N/A	185	N/A
R^2	0.180	N/A	0.669	N/A	0.159	N/A
Chi-square	35.16	N/A	146.67	N/A	33.65	N/A
Root mean square error	0.8481	N/A	0.6629	N/A	0.7399	N/A
P value	N/A	<.001	N/A	<.001	N/A	<.001

[a]DV: dependent variable.

[b]Standard errors in parentheses.

[c]N/A: not applicable.

The estimated coefficients were further compared across countries using a model-based chi-square comparison test with Bonferroni adjustments, as shown in Table 5. Multimedia Appendix 1 Table A4 further compares the coefficients across countries on the detailed PMT variables (*severity*, *vulnerability*, *self-efficacy*, and *response efficacy*).

Table 5. Comparison of coefficients across countries on the main variables.

Variables	US vs Kuwait		US vs South Korea		South Korea vs Kuwait	
	Chi-square[a]	P value	Chi-square	P value	Chi-square	P value
DV[b]: threat appraisal						
COVID-19 information sources	0.49	.48	3.19	.07	1.98	.16
COVID-19 social media	5.59	.02	4.98	.03	1.17	.28
DV: coping appraisal						
COVID-19 information sources	0.38	.54	2.33	.13	3.96	.047
COVID-19 social media	0.03	.87	0.02	.88	0.06	.80
DV: adherence						
Threat appraisal	0.27	.49	11.91	<.001	12.45	<.001
Coping appraisal	4.47	.049	1.66	.27	0.25	.62
Knowledge	0.24	.72	0.04	.81	0.01	.94

[a]Chi square values reported with Bonferroni adjustment.

[b]DV: dependent variable.

The first set of findings examines the *adherence* model in Tables 2 and 4. We first examined the effect of the *threat appraisal* variable on *adherence*. As shown in Tables 2 and 4, coefficients of *threat appraisal* were positive and statistically significant at P<.001 across the full sample (Table 2 column 3), the United States (P=.006; Table 4 column 1), and South Korea (P<.001; Table 4 column 2). This suggests that individuals who are threatened by the COVID-19 pandemic positively follow adherence guidance on social distancing. Further analysis of the *severity* coefficients in Multimedia Appendix 1 Table A3 shows positive and statistically significant influence on adherence (column 1: all sample P<.001; column 2: the United States P=.03; column 3: South Korea P<.001; column 4: Kuwait P<.001). However, as shown in Multimedia Appendix 1 Table A3, *vulnerability* was not statistically significant in any of the samples. This finding shows that *threat appraisal* was mainly led by *severity* in terms of *adherence* to COVID-19 social distancing regulations. As displayed in Table 5, comparing the

coefficients across countries, we found that the significant positive effect of *threat appraisal* on *adherence* shows that individuals who reside in the United States are more likely to follow adherence than South Korea, and those that reside in South Korea are more likely to follow adherence than those in Kuwait (*P*<.001).

Similarly, as displayed in Multimedia Appendix 1 Table A4, comparing the coefficients across countries, we found that the significant positive effect of *severity* on *adherence* shows that individuals who reside in the United States are more likely to follow adherence than South Korea (*P*=.02), and those in South Korea are more likely to follow adherence than those in Kuwait (*P*=.046). As shown in Multimedia Appendix 1 Table A4, in terms of *vulnerability*, the positive and significant effect showed that *vulnerability* was more influential on *adherence* for individuals that reside in South Korea than Kuwait (*P*=.07).

We subsequently examined the effect of *coping appraisal* on *adherence*. As shown in Tables 2 and 4, we found that the coefficient of *coping appraisal* was positive and statistically significant for all the models (Table 2 column 1: full sample, *P*<.001; Table 4 column 1: the United States, *P*=.002; Table 4 column 2: South Korea, *P*=.001; Table 4 column 3: Kuwait, *P*<.001). Further analysis in Multimedia Appendix 1 Table A3 showed that *self-efficacy* was positive and statistically significant in the full model (*P*=.02) and Kuwait only (*P*=.01). However, *response efficacy* was positive and statistically significant in all the models (column 1: full sample, *P*<.001; column 2: the United States, *P*<.001; column 3: South Korea, *P*<.001; column 4: Kuwait, *P*=.001). Interestingly, across countries, we found that, as displayed in Table 5, the significant positive impact of *coping appraisal* showed that those in the United States were more likely to follow adherence than those in Kuwait (*P*=.049). In terms of *self-efficacy* across countries, as shown in Multimedia Appendix 1 Table A4, we found the significant positive effect of *self-efficacy* on *adherence* showed that individuals who reside in the United States and South Korea are more likely to follow adherence than those in Kuwait (*P*<.001 and *P*=.04, respectively). As for *response efficacy* across countries, there was no significant comparative difference in this effect, as the comparative chi-square values were not significant, as shown in Multimedia Appendix 1 Table A4.

We then examined the effect of *knowledge* on *adherence*. As shown in Tables 2 and 4, the coefficients of *knowledge* were positive and statistically significant (Table 2 column 3: all sample, *P*<.001; Table 4 column 1: the United States, *P*=.01; Table 4 column 3: Kuwait *P*=.03) on *adherence* except for in South Korea. As shown in Table 5, comparisons of coefficients for *knowledge* across the three countries did not show statistically significant results.

The second set of findings examined the *threat appraisal* model in Tables 2 and 3. As shown in Table 2 column 2, *COVID-19 social media* coefficients were positive and statistically significant at *P*<.001 across the full sample and the United States (Table 3 column 4). This suggests that social media platforms have an impact on threat appraisal. However, we found no significance for *COVID-19 information sources*. Comparing coefficients across countries as shown in Table 5, we found that

the significant positive effect of *COVID-19 social media* on *threat appraisal* showed that individuals who reside in the United States were more likely to have higher *threat appraisal* than those in Kuwait (*P*=.02) and those in South Korea (*P*=.03).

Further analyzing the social media platform categories as shown in Multimedia Appendix 1 Table A5, in the United States, social network platforms such as Facebook and LinkedIn positively influenced threat appraisal (*P*=.03), and text-based or microblogging platforms such as Twitter and Whatsapp positively influenced threat appraisal in the whole sample (*P*<.001) and in the United States (*P*=.007). Some controls of the *threat appraisal* model in Tables 2 and 3 were significant. Females displayed higher *threat appraisal* than males in the full sample and in Kuwait (Table 2 column 2: *P*=.04 and Table 3 column 6: *P*=.02, respectively). In the United States, the lower the household income, the higher the *threat appraisal* (Table 3 column 4: *P*=.02).

The last set of findings examined the *coping appraisal* model in Tables 2 and 3. In contrast to the *threat appraisal* model, *COVID-19 social media* displayed no significance. However, coefficients for *COVID-19 information sources* were positive and significant for the whole sample (Table 2 column 1: *P*<.001), the United States (Table 3 column 1: *P*=.01), and South Korea (Table 3 column 2: *P*=.007). This finding suggests that using more information sources has an impact on the coping appraisal. Comparing coefficients across countries as shown in Table 5, we found that the significant positive effect of *COVID-19 information sources* on *coping appraisal* showed that individuals who reside in South Korea were more likely to have higher *coping appraisal* than those in Kuwait (*P*=.047).

Further analyzing the social media platform categories as shown in Multimedia Appendix 1 Table A5, social network platforms such as Facebook and LinkedIn negatively influenced coping appraisal on the whole sample (*P*=.001), and text-based or microblogging platforms such as Twitter and Whatsapp positively influenced coping appraisal in the whole sample (*P*<.001). Some controls of the *coping appraisal* model in Tables 2 and 3 were significant. Older individuals in the United States had lower *coping appraisal* (Table 3 column 1: *P*=.01). Compared to males, females had higher *coping appraisal* in the full model (Table 2 column 1: *P*=.005). Those with higher household income had higher *coping appraisal* in the full model (Table 2 column 1: *P*=.01).

Discussion

Principal Findings

In general, this study found that coping appraisal, threat appraisal, and knowledge positively influence adherence. Furthermore, using various COVID-19 information sources influences coping appraisal, and using social media for COVID-19 information influences threat appraisal. Tables 6 and 7 summarize the findings from this study (Multimedia Appendix 1 Table A6 displays the summary of findings on the specific PMT constructs and the social media platform constructs). In this section, we elaborate on the findings to provide useful policy and managerial insights.

Table 6. Summary of findings (part 1).

Variables	Coping appraisal				Threat appraisal				Findings
	All	US	South Korea	Kuwait	All	US	South Korea	Kuwait	
COVID-19 information sources	Pos[a]	Pos	Pos	NS[b]	NS	NS	NS	NS	H1[c]: Partially supported (supported for the whole sample, US, and South Korea). Using more information sources to get COVID-19 information positively influences coping appraisal. Information sources are more influential on coping appraisals in South Korea than Kuwait. No comparative difference in threat appraisal across countries.
COVID-19 social media	NS	NS	NS	NS	Pos	Pos	NS	NS	H2: Partially supported (supported for the whole sample and US). Using social media to get COVID-19 information positively influences threat appraisal. Social media is more influential on threat appraisal in US than Kuwait and South Korea. No comparative difference in coping appraisal across countries.

[a]Pos: positive association.

[b]NS: not significant.

[c]H: hypothesis.

Table 7. Summary of findings (part 2).

Variables	Adherence				Findings
	All	US	South Korea	Kuwait	
Threat appraisal	Pos[a]	Pos	Pos	NS[b]	H3[c]: Partially supported (supported for the whole sample, US, and South Korea). Threat appraisal positively influences social distancing adherence. Threat appraisal is more influential in US than in South Korea, and more in South Korea than in Kuwait.
Coping appraisal	Pos	Pos	Pos	Pos	H4: Supported. Coping appraisal positively influences social distancing adherence. Coping is more influential in US than Kuwait in terms of adherence.
Knowledge	Pos	Pos	NS	Pos	Knowledge positively influences social distancing adherence in the whole sample, US, and Kuwait. No comparative difference in results across countries.

[a]Pos: positive association.

[b]NS: not significant.

[c]H: hypothesis.

First, the study found that using more information sources positively influences the coping appraisal associated with the COVID-19 pandemic situation. Undoubtedly, this finding's importance relates to the increasing use and trade-off of information sources by the public. As much as people like to use many information sources, whether it is beneficial or not remains a question. Furthermore, citizens may be swayed by popular information sources such as the internet, TV, and newspapers. This finding highlights the benefits of using multiple information sources during the COVID-19 pandemic. This could be due to the high amount of misinformation and that the pandemic is ever-evolving. Therefore, attaining multiple information sources allows individuals to obtain more accurate information that helps with their coping appraisal. This finding is in line with a recent COVID-19 study that displayed an association between more sources of information and higher self-confidence to cope with COVID-19 in health care workers, yet the direction of the association was not confirmed in their study [22]. Thus, this study sheds light on this literature and

displays the importance of multiple sources of information to cope with the COVID-19 pandemic [48]. The study also found that information sources are more influential on the coping appraisal in South Korea than Kuwait, displaying that individuals in South Korea are more rationally using information sources to cope than individuals in Kuwait.

Second, this study found that using social media to get COVID-19 information positively influences threat appraisal. This is an important finding to validate the increasing trustworthiness of social media on people's decisions and the role of social media in this pandemic. Social media is one of the main channels used to provide updated COVID-19 information [49]. This study's findings are consistent with previous studies [50] that showed that the frequency of social media was associated with high odds of anxiety [51]. This association might be because, throughout the COVID-19 pandemic, many false reports and misinformation bombarded social media, which resulted in a lot of confusion and anxiety [50]. Furthermore, many individuals use social media to express

their feelings of worry, anxiety, nervousness, and fear, which is contagious in social networks [52]. As we stated earlier, the social media role in people's minds during the pandemic is not free from harmful aspects such as increased anxiety, sleep disturbance, suicidal thoughts[19], misinformation, and distress [20,21]. Thus, it is interesting to note the positive aspects such as social media's role in adherence decisions. As for country comparisons, we found that social media is more influential on threat appraisal in the United States than in South Korea and Kuwait. A possible reason for this finding is that social media information in the United States might be more fearful or driven with more riddles (ie, whether to trust Twitter postings or not) compared to South Korea and Kuwait.

Third, threat appraisal positively influences social distancing adherence. One of the primary emotional responses during a pandemic is fear [53]. This fear is a defensive system to combat ecological threats [54]. A meta-analysis found that targeting fears can be useful in situations, such that appealing to fear leads people to change their behavior if they feel capable of dealing with the threat but leads to defensive reactions when they feel helpless to act [55]. This study sheds light on this stream of research by displaying that fear or threat appraisal was useful in the COVID-19 pandemic since it changes an individual's behavior and influences their social distancing adherence.

Based on PMT, a fear appeal is a cognitive assessment that prepares individuals against the severity of a threatening event. People consider the pros and cons, and the probability of the event occurring to develop a response. The recommended response in the COVID-19 context is adherence. The cognitive assessment enhances the fear appeal and subsequent motivation to protect oneself—without which the recommended action would not be effective across citizens. Besides, there is a more substantial threat appraisal for social distance adherence in the United States than in South Korea and Kuwait. This follows through with the social media country comparison finding; in addition, the finding shows that threat appraisal in the United States is more rational than that of South Korea and Kuwait.

Fourth, coping appraisal positively influences social distancing adherence, and this effect is more influential in the United States than Kuwait, again displaying that individuals in the United States are more pragmatic in the coping and adherence process. One of the main factors that have been stressed upon during the COVID-19 pandemic is how to cope with the current situation [15]. Studies have shown that problem-focused coping is associated with better adherence to health-related behaviors and a higher sense of control [56,57]. The findings support these perspectives and suggest, as per PMT, that the preventive actions will be preferred in a high threat situation when both the self-efficacy and the efficacy of the adherence plans are high. This finding agrees with this research stream and displays the importance of coping appraisal during the COVID-19 pandemic to adhere to social distancing recommendations.

These findings partially explain the variations observed in the adherence outcomes, which can be explained due to the variations observed in the threat and coping appraisals. Finally, as an additional outcome, we found that knowledge positively influences social distancing adherence in the whole sample, the United States, and Kuwait. There were no comparative difference in results across countries. This sheds light on the importance of correct knowledge during this pandemic to be able to adhere to social distancing recommendations. Association between adherence and better knowledge of the disease is consistent with recent findings and provides evidence for using proper interventions of proper communication related to the disease on improving adherence [56,58].

Practice and Policy Implications

A set of implications and recommendations for public health officials and policy makers can be drawn from this study. This study points to the gap in the responsible behavior of individuals regarding adherence. The current government efforts to mitigate and the expectation that everything should be back to normal have differing consequences. Existing work suggests that policy recommendations' efficacy and outcomes depend on the individuals' beliefs and subsequent actions [59,60]. The first step in this process is the firm belief of whether the recommended action will mitigate the threat or manage the fearful situation.

Thus, the key to minimizing rejection and maximizing acceptance of recommendation is reliant on the messages and information provided to citizens. Often, the media outlets and information sources are left to craft their messages independently or freely without the policy makers' recommended guidelines. There needs to be policy guidelines with careful attention about what messages to give and how to give them, so that it does not instill too much fear nor allow citizens to be too careless (ie, messages and information need to have a balance to involve appropriate threat and coping appraisals in citizens' minds).

Limitations

This study examines factors that influence citizens adhering to social distancing at a point in time. However, the citizen might go back and forth in the adherence process, and the threat appraisal, coping appraisal, knowledge level, and sources of information may change over time. This is a limitation of this study, as the data set used is a cross-sectional survey. Future studies could examine how a citizen's coping appraisal, threat appraisal, and adherence changes over time.

In addition, using the random sampling process in the United States, Kuwait, and South Korea samples may include fewer familiar respondents to the study context. In particular, the questionnaire was only online. Therefore, respondents were all users of the internet. The study does not examine noninternet users, which could have differential impacts. Thus, the generalization of the sample to a uniform national culture characteristic is a limitation of this study. Future studies could conduct online and offline surveys, and examine the difference in threat and coping appraisal in terms of adherence behavior.

The findings should be taken with care due to the sample being representative of three countries; thus, future studies could expand to more countries to examine even more cultural differences. Another limitation is the small sample size for South Korea, limiting the generalizability of the results.

Conclusions

As the COVID-19 global pandemic continues to grow and governmental restrictions are ongoing, it is critical to understand people's frustration to reduce panic and promote social distancing to control the pandemic. This study points to a threat and coping appraisal mechanism that may clarify the adherence variations. This study also highlights that social media has an impact on the threat appraisal of the COVID-19 pandemic. Furthermore, the intensity of information sources used to attain COVID-19 information impacts the coping appraisal for the pandemic.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire and detailed analysis.
[DOCX File , 129 KB - jmir_v22i11e23019_app1.docx]

References

1. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-05-01]
2. Coronavirus disease (COVID-19). World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses [accessed 2020-05-10]
3. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA 2020 Apr 14;323(14):1406-1407 [FREE Full text] [doi: 10.1001/jama.2020.2565] [Medline: 32083643]
4. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
5. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. Asian Pac J Allergy Immunol 2020 Mar;38(1):1-9 [FREE Full text] [doi: 10.12932/AP-200220-0772] [Medline: 32105090]
6. Fisher D, Heymann D. BMC Med 2020 Feb 28;18(1):57 [FREE Full text] [doi: 10.1186/s12916-020-01533-w] [Medline: 32106852]
7. Ahorsu DK, Imani V, Lin C, Timpka T, Broström A, Updegraff JA, et al. Associations between fear of COVID-19, mental health, and preventive behaviours across pregnant women and husbands: an actor-partner interdependence modelling. Int J Ment Health Addict 2020 Jun 11:1-15 [FREE Full text] [doi: 10.1007/s11469-020-00340-x] [Medline: 32837427]
8. Chang K, Strong C, Pakpour AH, Griffiths MD, Lin C. Factors related to preventive COVID-19 infection behaviors among people with mental illness. J Formos Med Assoc 2020 Jul 29:S0929-6646(20)30344-2 [FREE Full text] [doi: 10.1016/j.jfma.2020.07.032] [Medline: 32773260]
9. Lin C, Imani V, Majd NR, Ghasemi Z, Griffiths MD, Hamilton K, et al. Using an integrated social cognition model to predict COVID-19 preventive behaviours. Br J Health Psychol 2020 Nov;25(4):981-1005 [FREE Full text] [doi: 10.1111/bjhp.12465] [Medline: 32780891]
10. Lin M, Cheng Y. Policy actions to alleviate psychosocial impacts of COVID-19 pandemic: experiences from Taiwan. Soc Health Behav 2020;3(2):72. [doi: 10.4103/shb.shb_18_20]
11. Usman N, Mamun M, Ullah I. COVID-19 infection risk in pakistani health-care workers: the cost-effective safety measures for developing countries. Soc Health Behav 2020;3(3):75. [doi: 10.4103/shb.shb_26_20]
12. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
13. Rieger M. To wear or not to wear? Factors influencing wearing face masks in Germany during the COVID-19 pandemic. Soc Health Behav 2020;3(2):50. [doi: 10.4103/shb.shb_23_20]
14. Rieger M. Triggering altruism increases the willingness to get vaccinated against COVID-19. Soc Health Behav 2020;3(3):78. [doi: 10.4103/shb.shb_39_20]
15. Stress and coping. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/stress-coping/index.html [accessed 2020-06-08]
16. Satici B, Gocet-Tekin E, Deniz ME, Satici SA. Adaptation of the fear of COVID-19 scale: its association with psychological distress and life satisfaction in Turkey. Int J Ment Health Addict 2020 May 08:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00294-0] [Medline: 32395095]
17. Montemurro N. The emotional impact of COVID-19: from medical staff to common people. Brain Behav Immun 2020 Jul;87:23-24 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.032] [Medline: 32240766]

XSL•FO
RenderX

18.   Lin C. Social reaction toward the 2019 novel coronavirus (COVID-19). Soc Health Behav 2020;3(1):1. [doi: 10.4103/shb.shb_11_20]

19.   Li D, Ko N, Chen Y, Wang P, Chang Y, Yen C, et al. COVID-19-related factors associated with sleep disturbance and suicidal thoughts among the Taiwanese public: a Facebook survey. Int J Environ Res Public Health 2020 Jun 22;17(12):4479 [FREE Full text] [doi: 10.3390/ijerph17124479] [Medline: 32580433]

20.   Olatunji O, Ayandele O, Ashirudeen D, Olaniru O. "Infodemic" in a pandemic: COVID-19 conspiracy theories in an african country. Soc Health Behav 2020;3(4):152. [doi: 10.4103/shb.shb_43_20]

21.   Lin C, Broström A, Griffiths MD, Pakpour AH. Investigating mediated effects of fear of COVID-19 and COVID-19 misunderstanding in the association between problematic social media use, psychological distress, and insomnia. Internet Interv 2020 Sep;21:100345 [FREE Full text] [doi: 10.1016/j.invent.2020.100345] [Medline: 32868992]

22.   Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: Facebook survey study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]

23.   Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]

24.   Ornell F, Schuch JB, Sordi AO, Kessler FHP. "Pandemic fear" and COVID-19: mental health burden and strategies. Braz J Psychiatry 2020;42(3):232-235 [FREE Full text] [doi: 10.1590/1516-4446-2020-0008] [Medline: 32267343]

25.   Aronson E, Akert RM, Wilson TD. Sozialpsychologie. Hallbergmoos, Germany: Pearson Deutschland GmbH; 2010.

26.   Baron RS, Vandello JA, Brunsman B. The forgotten variable in conformity research: impact of task importance on social influence. J Pers Soc Psychol 1996;71(5):915-927. [doi: 10.1037/0022-3514.71.5.915]

27.   Lucas T, Alexander S, Firestone IJ, Baltes BB. Self‐efficacy and independence from social influence: discovery of an efficacy–difficulty effect. Soc Influence 2006 Mar;1(1):58-80. [doi: 10.1080/15534510500291662]

28.   Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol 2004;55:591-621. [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]

29.   Bhagavathula AS, Aldhaleei WA, Rahmani J, Mahabadi MA, Bandari DK. Knowledge and perceptions of COVID-19 among health care workers: cross-sectional study. JMIR Public Health Surveill 2020 Apr 30;6(2):e19160 [FREE Full text] [doi: 10.2196/19160] [Medline: 32320381]

30.   Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]

31.   Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]

32.   Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]

33.   Timpka T, Spreco A, Gursky E, Eriksson O, Dahlström Ö, Strömgren M, et al. Intentions to perform non-pharmaceutical protective behaviors during influenza outbreaks in Sweden: a cross-sectional study following a mass vaccination campaign. PLoS One 2014;9(3):e91060 [FREE Full text] [doi: 10.1371/journal.pone.0091060] [Medline: 24608557]

34.   Floyd DL, Prentice-Dunn S, Rogers RW. A meta-analysis of research on protection motivation theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]

35.   Milne S, Sheeran P, Orbell S. Prediction and intervention in health-related behavior: a meta-analytic review of protection motivation theory. J Appl Social Pyschol 2000 Jan;30(1):106-143. [doi: 10.1111/j.1559-1816.2000.tb02308.x]

36.   Lee SA. Coronavirus Anxiety Scale: a brief mental health screener for COVID-19 related anxiety. Death Stud 2020;44(7):393-401. [doi: 10.1080/07481187.2020.1748481] [Medline: 32299304]

37.   Lee SA. How much "Thinking" about COVID-19 is clinically dysfunctional? Brain Behav Immun 2020 Jul;87:97-98 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.067] [Medline: 32353520]

38.   Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The fear of COVID-19 scale: development and initial validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]

39.   Chang K, Hou W, Pakpour AH, Lin C, Griffiths MD. Psychometric testing of three COVID-19-related scales among people with mental illness. Int J Ment Health Addict 2020 Jul 11:1-13 [FREE Full text] [doi: 10.1007/s11469-020-00361-6] [Medline: 32837442]

40.   Rahaei Z, Ghofranipour F, Morowatisharifabad MA, Mohammadi E. Psychometric properties of a protection motivation theory questionnaire used for cancer early detection. J Sch Public Health Institute Public Health Res 2015;12(3):69-79.

41.   Xiao H, Li S, Chen X, Yu B, Gao M, Yan H, et al. Protection motivation theory in predicting intention to engage in protective behaviors against schistosomiasis among middle school students in rural China. PLoS Negl Trop Dis 2014 Oct;8(10):e3246 [FREE Full text] [doi: 10.1371/journal.pntd.0003246] [Medline: 25329829]

42. Yan Y, Jacques-Tiura AJ, Chen X, Xie N, Chen J, Yang N, et al. Application of the protection motivation theory in predicting cigarette smoking among adolescents in China. Addict Behav 2014 Jan;39(1):181-188 [FREE Full text] [doi: 10.1016/j.addbeh.2013.09.027] [Medline: 24157424]

43. Macdonell K, Chen X, Yan Y, Li F, Gong J, Sun H, et al. A protection motivation theory-based scale for tobacco research among Chinese youth. J Addict Res Ther 2013 Jul 08;4:154 [FREE Full text] [doi: 10.4172/2155-6105.1000154] [Medline: 24478933]

44. Hassani L, Dehdari T, Hajizadeh E, Shojaeizadeh D, Abedini M, Nedjat S. Development of an instrument based on the protection motivation theory to measure factors influencing women's intention to first pap test practice. Asian Pac J Cancer Prev 2014;15(3):1227-1232 [FREE Full text] [doi: 10.7314/apjcp.2014.15.3.1227] [Medline: 24606445]

45. Xiao H, Peng M, Yan H, Gao M, Li J, Yu B, et al. An instrument based on protection motivation theory to predict Chinese adolescents' intention to engage in protective behaviors against schistosomiasis. Glob Health Res Policy 2016;1:15 [FREE Full text] [doi: 10.1186/s41256-016-0015-6] [Medline: 29202064]

46. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]

47. Frequently asked questions. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/faq.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Fchildren-faq.html [accessed 2020-05-02]

48. Al-Hasan A, Yim D, Khuntia J. Citizens' adherence to COVID-19 mitigation recommendations by the government: a three-country comparative evaluation using online cross-sectional survey data. J Med Internet Res 2020 Aug 11;22(8):e20634 [FREE Full text] [doi: 10.2196/20634] [Medline: 32716896]

49. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]

50. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

51. Neria Y, Sullivan GM. Understanding the mental health effects of indirect exposure to mass trauma through the media. JAMA 2011 Sep 28;306(12):1374-1375 [FREE Full text] [doi: 10.1001/jama.2011.1358] [Medline: 21903818]

52. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]

53. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

54. Mobbs D, Hagan CC, Dalgleish T, Silston B, Prévost C. The ecology of human fear: survival optimization and the nervous system. Front Neurosci 2015;9:55. [doi: 10.3389/fnins.2015.00055] [Medline: 25852451]

55. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]

56. Vélez-Vélez E, Bosch RJ. Illness perception, coping and adherence to treatment among patients with chronic kidney disease. J Adv Nurs 2016 Apr;72(4):849-863. [doi: 10.1111/jan.12873] [Medline: 26689295]

57. Ransom S, Jacobsen PB, Schmidt JE, Andrykowski MA. Relationship of problem-focused coping strategies to changes in quality of life following treatment for early stage breast cancer. J Pain Symptom Manage 2005 Sep;30(3):243-253 [FREE Full text] [doi: 10.1016/j.jpainsymman.2005.03.013] [Medline: 16183008]

58. Parvan K, Hasankhani H, Seyyedrasooli A, Riahi SM, Ghorbani M. The effect of two educational methods on knowledge and adherence to treatment in hemodialysis patients: clinical trial. J Caring Sci 2015 Mar;4(1):83-93 [FREE Full text] [doi: 10.5681/jcs.2015.009] [Medline: 25821762]

59. Mukhtar S. Mental health and emotional impact of COVID-19: applying Health Belief Model for medical staff to general public of Pakistan. Brain Behav Immun 2020 Jul;87:28-29 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.012] [Medline: 32283289]

60. Guidry JPD, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the Health Belief Model to analyze Instagram posts about Zika for public health communications. Emerg Infect Dis 2019 Jan;25(1):179-180. [doi: 10.3201/eid2501.180824] [Medline: 30561302]

Abbreviations

PMT: protection motivation theory
SUR: seemingly unrelated regression

XSL·FO
RenderX

<u>Original Paper</u>

Status of Institutional Review Board Meetings Conducted Through Web Conference Systems in Japanese National University Hospitals During the COVID-19 Pandemic: Questionnaire Study

Kenta Yagi[1], PhD; Kazuki Maeda[1], BA; Satoshi Sakaguchi[1], MD, PhD; Masayuki Chuma[1], PhD; Yasutaka Sato[1], BA; Chikako Kane[1], PhD; Akiyo Akaishi[1], RNU; Keisuke Ishizawa[2,3], PhD; Hiroaki Yanagawa[1], MD, PhD

[1]Clinical Research Center for Developmental Therapeutics, Tokushima University Hospital, Tokushima, Japan

[2]Department of Clinical Pharmacology and Therapeutics, Tokushima University Graduate School of Biomedical Sciences, Kuramoto-cho, Tokushima, Japan

[3]Department of Pharmacy, Tokushima University Hospital, Kuramoto-cho, Tokushima, Japan

Corresponding Author:
Kenta Yagi, PhD
Clinical Research Center for Developmental Therapeutics
Tokushima University Hospital
2-50-1 Kuramoto-cho
Tokushima, 770-8503
Japan
Phone: 81 88 615 8512
Fax: 81 88 633 9295
Email: yagi.kenta@tokushima-u.ac.jp

Abstract

Background: With the global proliferation of the novel COVID-19 disease, conventionally conducting institutional review board (IRB) meetings has become a difficult task. Amid concerns about the suspension of drug development due to delays within IRBs, it has been suggested that IRB meetings should be temporarily conducted via the internet.

Objective: This study aimed to elucidate the current status of IRB meetings conducted through web conference systems.

Methods: A survey on conducting IRB meetings through web conference systems was administered to Japanese national university hospitals. Respondents were in charge of operating IRB offices at different universities. This study was not a randomized controlled trial.

Results: The survey was performed at 42 facilities between the end of May and early June, 2020, immediately after the state of emergency was lifted in Japan. The survey yielded a response rate of 74% (31/42). Additionally, while 68% (21/31) of facilities introduced web conference systems for IRB meetings, 13% (4/31) of the surveyed facilities postponed IRB meetings. Therefore, we conducted a further survey of 21 facilities that implemented web conference systems for IRB meetings. According to 71% (15/21) of the respondents, there was no financial burden for implementing these systems, as they were free of charge. In 90% (19/21) of the facilities, IRB meetings through web conference systems were already being conducted with personal electronic devices. Furthermore, in 48% (10/21) of facilities, a web conference system was used in conjunction with face-to-face meetings.

Conclusions: Due to the COVID-19 pandemic, the number of reviews in clinical trial core hospitals has decreased. This suggests that the development of pharmaceuticals has stagnated because of COVID-19. According to 71% (15/21) of the respondents who conducted IRB meetings through web conference systems, the cost of introducing such meetings was US $0, showing a negligible financial burden. Moreover, it was shown that online deliberations could be carried out in the same manner as face-to-face meetings, as 86% (18/21) of facilities stated that the number of comments made by board members did not change. To improve the quality of IRB meetings conducted through web conference systems, it is necessary to further examine camera use and the content displayed on members' screens during meetings. Further examination of all members who use web conference systems is required. Our measures for addressing the requests and problems identified in our study could potentially be considered protocols for future IRB meetings, when the COVID-19 pandemic has passed and face-to-face meetings are possible again. This study also highlights the importance of developing web conference systems for IRB meetings to respond to future unforeseen pandemics.

XSL•FO
RenderX

Introduction

In 2020, the novel COVID-19 disease became a pandemic [1], causing several changes to people's lifestyles [2]. In Japan, many individuals were infected, prompting the declaration of a state of emergency [3] for the entire country. As a result, people have avoided the "3 Cs"—closed spaces, crowds, and contact. Organizations, including institutional review boards (IRBs), have also adopted this practice, making it difficult to conduct regular face-to-face meetings. Without hospital IRBs, clinical trials cannot be conducted [4,5]. Therefore, concerns about the stagnation of drug development have been raised. Japanese ordinances indicate that IRBs should operate in all medical research institutes [5]. Within this context, the Ministry of Health, Labor, and Welfare [6] issued a notice titled "On handling of reviews by the Institutional Review Board regarding clinical trials related to the novel coronavirus," which advised that IRB meetings were allowed to be conducted via the internet. Due to the declaration of a state of emergency and university regulations, it has become difficult to conduct face-to-face IRB meetings in a group setting at Tokushima University hospital. Therefore, we attempted to conduct IRB meetings through a web conference system at our hospital, starting on May, 2020.

Thanks to advancements in science and technology, online communication has become normal in human life due to the widespread use of the internet [7,8]. As a result, web conference systems are being adopted in places of work, rapidly gaining popularity, and expanding their market scale. Due to the COVID-19 pandemic, the demand for web communication has rapidly increased [9,10]. Web conferencing systems are also being used more in health care [11-15]. However, online communication requires an operating environment with internet connectivity and compatible devices, which can be costly. Moreover, internet meetings are relatively new, and various problems during their implementation have been reported [16-18]. Unfortunately, there are no clear, useful countermeasures for these issues. There may also be problems within IRBs, such as proper deliberation upon holding a web-based IRB meeting and the possibility of information leakage. The actual situation regarding web-based IRB meetings remains unclear because the problems surrounding such meetings have yet to be thoroughly examined. As a result, the necessary steps for conducting a successful web-based IRB meeting remains unclear. In addition, since the social environment on the internet and the laws and regulations on IRBs differ across countries, it is expected that individual, country-specific countermeasures are required. Based on the web conference system already implemented in our hospital for IRB meetings, we surveyed external members who wanted to continue having meetings through the web conference system.

Therefore, conducting IRB meetings through web conference systems may continue, even after the COVID-19 pandemic resolves. However, it is necessary to ensure the efficiency of such meetings while maintaining proper deliberations. As such,

we aimed to clarify the current status of IRB meetings conducted through web conference systems and elucidate problems related to such meetings across Japan.

Methods

Questionnaire Administered to Japanese National University Hospitals

With the cooperation of Topic Group 1, a subcommittee of the National University Hospital Clinical Research Promotion Initiative, we administered a closed survey to participating facilities regarding IRB meetings conducted through web conference systems. The group is composed of national university hospitals. The respondents were in charge of operating the IRB office of each university, and agreements were based on the questionnaire's answers. To understand the situation during the declaration of a state of emergency (ie, April to May, 2020), the survey was conducted immediately after its announcement (ie, end of May to early June, 2020) in Japan. The questionnaire used for the survey was based on the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [19] and made using Microsoft Forms. We delivered the survey via email, and the respondents answered using a computer. To avoid largely varying opinions, the respondents were asked to select answers from predefined options. The questionnaire consisted of 39 questions and 3 pages. The third and subsequent questions were provided in a table, and they targeted facilities that had conducted IRB meetings through a web conference system. The questions are listed in Table S1 in Multimedia Appendix 1. We confirmed that there were no multiple responses from the same respondents, as there was a form for noting respondents' university affiliations. The response rate was 74% (31/42). Although a structure was in place for contacting respondents when answers were incomplete, it was not used because no respondents were excluded from the analysis due to missing data. This survey did not require a review by the IRB because there were no items regarding personal information and this study was not a randomized controlled trial. No incentives were generated for participating in this survey.

Analysis Methods

Based on the number of people who contracted COVID-19, we established endemic and nonendemic areas. As of May 31, 2020, a COVID-19 epidemic area is defined as a region with ≥500 COVID-19–positive cases. Additionally, the data regarding the COVID-19 epidemic were tallied in the period following Japan's first confirmed COVID-19 case in February, 2020. With regard to items related to cost, currencies were converted from yen and stated as US dollars (¥100=US $1). In this study, external members were defined as members who were not Good Clinical Practice experts and members who were not otherwise affiliated with the involved parties.

In Japan, core hospitals are designated as sites that lead clinical trials. Since IRB operation methods and sizes may be different for core hospitals that lead clinical trials for other facilities, we collected data on the IRB environment of other facilities separately. In addition, although the participation of external committee members is required to hold an IRB meeting, the "3Cs" must be avoided for IRB participation. Therefore, the number of external committee members may be a factor in conducting IRB meetings through web conference systems. As such, data on external committee members were also separately collected.

Statistical Analyses

The composition of IRB members in Japanese university hospitals (Figure 1) and the number of new registrations to IRBs in Japanese university hospitals before and during the COVID-19 pandemic (Figure 2) were analyzed using the Student t test. The digitization of review documents in Japanese university hospitals (Figure 3) were compared using the Chi-square test. Data were shown as means with standard deviations or n (%). A P value of <.05 was considered statistically significant for all analyses.

Figure 1. Composition of institutional review board members in Japanese university hospitals.
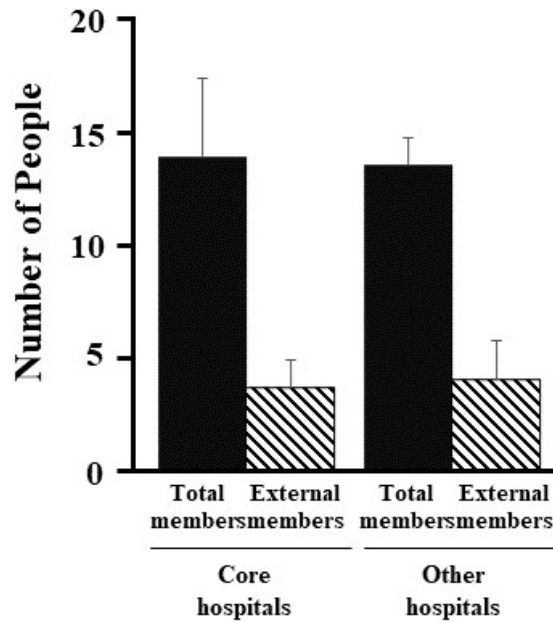
Figure 2. Comparison between the number of new registrations to institutional review boards at university hospitals in Japan before and after the COVID-19 pandemic.
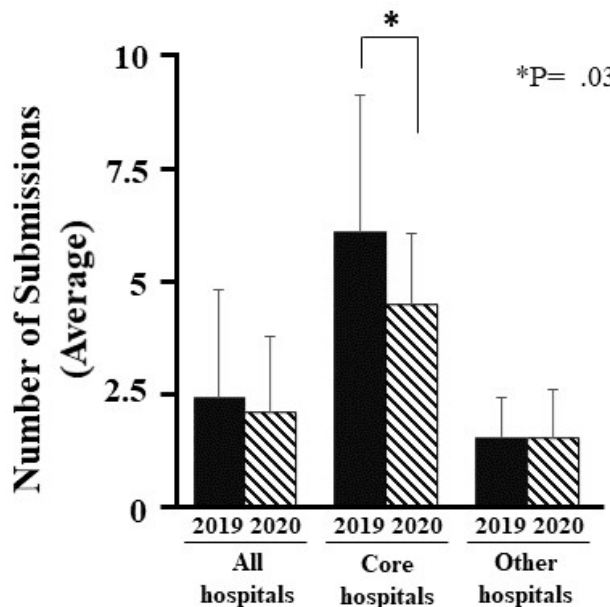
Figure 3. Digitization of review documents in Japanese university hospitals.
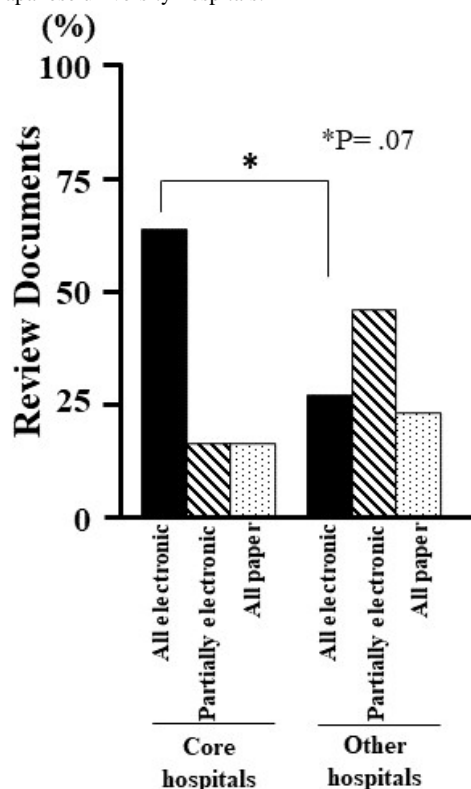
Results

Status of Facilities Before the COVID-19 Pandemic

Of the 31 facilities that responded to the survey, 6 (19%) were clinical trial core hospitals [20]. IRB meetings were held monthly at all facilities. In addition, there was no difference in the total number of IRB members and external members based on whether the facility was a clinical research core hospital (Figure 1).

For core hospitals, all review materials were electronic in 67% (4/6) of facilities, partially electronic in 17% (1/6) of facilities, and all on paper in 17% (1/6) of facilities. For other hospitals, all review materials were electronic in 28% (7/25) of facilities, partially electronic in 48% (12/25) of facilities, and all on paper in 24% (6/25) of facilities. There was no significant difference in the number of facilities that digitized all review materials between core hospitals and other facilities (P=.07).

With regard to participation in IRB meetings, external members spent a median of US $20 per 40 minutes. It was also revealed that 16% (5/31) of respondents participated in meetings for 60 minutes or longer (Figures 2 and 4).

Figure 4. Cost required for external committee members to get to institutional review board meeting sites at university hospitals in Japan.
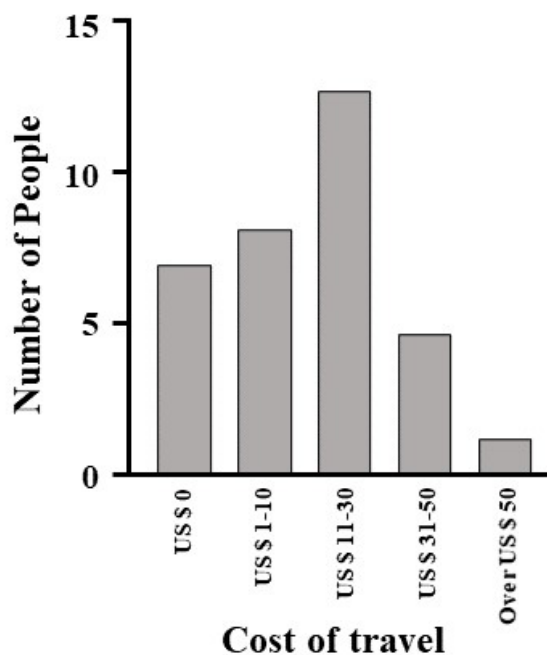
Strategies or Practices During the COVID-19 Pandemic

In the Strategies or Practices During COVID-19 section of the questionnaire, we examined the relationship between the COVID-19 pandemic and the introduction of web-based IRB meetings (Table 1). Of the 31 IRB offices, 11 (35%) worked remotely due to COVID-19, while 4 (13%) opted to postpone IRB meetings. About 68% (21/31) of the facilities introduced web conference systems for IRB meetings. However, 13% (4/31) of the universities could not hold IRB meetings due to reasons unrelated to the COVID-19 pandemic and 20% (1/5) were unable to do so due to being located in a COVID-19 epidemic area. The rate of introducing web conference systems for IRB meetings was 60% (3/5) for facilities located in a COVID-19 epidemic area.

Web conference systems were also used in 81% (17/21) of facilities that introduced web-based IRB meetings (Table 1). The rate of introducing web conference systems for IRB meetings was 68% (21/31) for all hospitals, 71% (17/24) for facilities that digitized review materials, and 57% (4/7) for facilities that did not.

Table 1. Activity and preparation regarding web-based institutional review board meetings in university hospitals during the COVID-19 pandemic.

Activity	IRB[a] member, n (%)
IRB office is conducting remote work	
Yes	11/31 (35)
No	20/31 (65)
IRB meetings are cancelled or suspended	
Yes	4/31 (13)
No	27/31 (87)
IRB meetings are conducted using web conferencing tools	
Yes	21/31 (68)
No	10/31 (32)

[a]IRB: institutional review board.

A further investigation was conducted on the implementation of web conference systems in facilities that held web-based IRB meetings (Table 2). Of the 21 facilities that conducted IRB meetings through web conference systems, 15 (71%) stated that the introduction cost was US $0, 19 (90%) held meetings using personal electronic devices, and all facilities kept recordings of these IRB meetings as minutes. However, the most popular web conference system, WebEX, was used in 38% (8/21) of facilities, and security measures that were implemented varied across facilities.

We also questioned whether the facilities that adopted web conference systems for IRB meetings were able to deliberate properly (Table 3). Of the 21 facilities, 15 (71%) stated that cameras were used during IRB meetings, and of those 15, 10 (67%) conducted IRB meetings with a combination of web

conference systems and face-to-face meetings. The display on members' computer screens during deliberation also varied according to the facility (ie, faces of other members, review materials, etc). Moreover, compared to in-person IRB meetings conducted before the COVID-19 pandemic, the discussion time

for IRB meetings conducted with web conference systems during the COVID-19 pandemic decreased in 38% (8/21) of facilities, but the number of member comments decreased in 10% (2/21) of the facilities.

Table 2. Practical aspects regarding the management of institutional review board meetings conducted through web conference systems.

IRB[a] meeting characteristics	IRB member, n (%)
Cost of hosting an IRB meeting using a web conference system (US $)	
0	15/21 (71)
100-500	2/21 (10)
501-1000	1/21 (5)
Unknown	3/21 (14)
Confirmation of attendance	
Using the list of participants in the web conferencing system	2/21(10)
Confirmation of attendance through live attendance	11/21 (52)
Voice confirmation	7/21(33)
Checklist of participants attending the web conference and confirmation through live attendance	1/21 (5)
Ownership of devices used	
Own devices	19/21 (90)
Provided by the committee, if needed	1/21 (5)
Distributed by the committee	1/21 (5)
Information for keeping track of IRB meetings[b]	
Note in the logbooks that the meeting was held on the web	21/21 (100)
Description of the system in use	5/21 (24)
Note on security policy	3/21 (14)
Note that there was enough time for discussion	3/21 (14)
Location of each committee member's place of participation	1/21 (5)
Web conference system used for IRB meeting	
Google Meet	2/21 (10)
Skype	2/21 (10)
Teams	2/21 (10)
WebEX	8/21 (38)
Zoom	7/21 (33)
Security policy	
Requested to be considerate of the surrounding environment during committee meetings	11/21 (52)
Use of the campus network	6/21 (29)
Installation of security software	6/21 (29)
Video recording is not allowed.	2/21(10)
Warning about the handling of the URLs distributed for conference participation	1/21 (5)
Entering a password to join a web conference	2/21(10)
Management and tracking of participant accounts	3/21 (14)
Updated operating system software	1/21 (5)

[a]IRB: institutional review board.

[b]Multiple options could be selected.

Table 3. Influence of the introduction of web conference systems for institutional review board meetings on the review process.

Activity	IRB[a] member, n (%)
Using video in web-based IRB meeting	
All members	15/21 (71)
According to member's choice	6/21 (29)
Members' information displayed on screen during the meeting	
Committee members on screen through camera	6/21 (29)
Review materials and participating committee members	14/21 (66)
Participant's choice	1/21 (5)
Members Using the Web Conferencing System	
All members	11/21 (52)
External members only	3/21 (14)
Participant's choice	7/21 (33)
Confirmation of decisions made	
Using the tools provided by the web conference system	1/21 (5)
All committee members are involved prior to decision making	1/21 (5)
Raising of hands	1/21 (5)
Verbal confirmation	17/21 (81)
Making use of the review table	1/21 (5)

[a]IRB: institutional review board.

Comparison of IRB Meetings Conducted Before and During the COVID-19 Pandemic

The number of new clinical trial reviews before and during the COVID-19 pandemic (ie, February to May, 2019 and February to May, 2020, respectively) was analyzed to investigate the impact of the COVID-19 pandemic on the conduction of clinical trials. Following the onset of the pandemic, the number of reviews conducted at clinical trial core hospitals decreased (Figure 2). No facility shortened the discussion time by more than 30 minutes, and the number of comments from committee members remained consistent in 86% (18/21) of the facilities (Table 4).

Table 4. Comparison of the discussion characteristics of institutional review board meetings conducted through web conference systems during the COVID-19 epidemic and those of in-person meetings conducted before the COVID-19 epidemic.

Discussion characteristics	IRB[a] member, n (%)
Discussion length compared to in-person meetings	
Shorter by <30 minutes	8/21 (38)
Longer by <30 minutes	2/21 (10)
Longer by >30 minutes	1/21 (5)
Same as before	10/21 (48)
Number of comments made by committee members compared to in-person meetings	
More than 1.5 times	1/21 (5)
Less than half	2/21 (10)
Same as before	18/21 (86)

[a]IRB: institutional review board.

Discussion

Principal Results

This study clarifies the problems that need to be solved to improve IRB meetings conducted through web conference systems. This was done by collecting information on the implementation of web conference systems for IRB meetings in Japan's national university hospitals.

In Japan, the declaration of a state of emergency [3] prompted the need to reexamine the way IRB meetings should be held. Along with this, the Ministry of Health, Labor, and Welfare

issued a notice [5], tentatively allowing IRB meetings to be held in a nonassembly form. In this study, we carried out a questionnaire survey among national university hospitals in the country after the official notice to understand the problems concerning IRB meetings conducted through web conference systems. In addition, we have identified the problems that need to be addressed to improve the quality of such IRB meetings.

After the onset of the COVID-19 pandemic, the number of reviews in clinical trial core hospitals has decreased compared to those before the pandemic. However, there were no significant changes in the number of reviews in other hospitals (Figure 2). This may have been influenced by the fact that 50% (3/6) of the clinical trial core hospitals were located in COVID-19 hotspots and that the number of applications per month at other hospitals was already low (ie, 1.5 applications on average) before the pandemic. However, since the number of reviews has decreased in clinical trial core hospitals, which oversee several clinical trials, there is growing concern about the stagnation of pharmaceutical development due to COVID-19.

Therefore, it is important to advance developments in IRB meetings conducted through web conference systems to prepare for future unforeseen pandemics that follow the COVID-19 pandemic.

Conducting IRB meetings through web conference systems allows people to avoid overcrowding, prevent infection, participate regardless of location, and reduce the time and cost required to commute to IRB meeting venues (Figures 4 and 5). Therefore, continuing the implementation web conference systems for IRB meetings may increase the number of attendees. Moreover, in the current IRB format, ensuring the quality of external members is difficult, especially in provincial cities, as these members would have to commute from neighboring areas. If the quality of IRBs can be guaranteed, even when using a web conference system, more advanced discussions will be possible. Web conference systems allow for the easy invitation of people who have deep insights, but are unable to join regular meetings due to the distance of IRB meeting venues.

Figure 5. Time required for external committee members to get to institutional review board meeting sites at university hospitals in Japan.
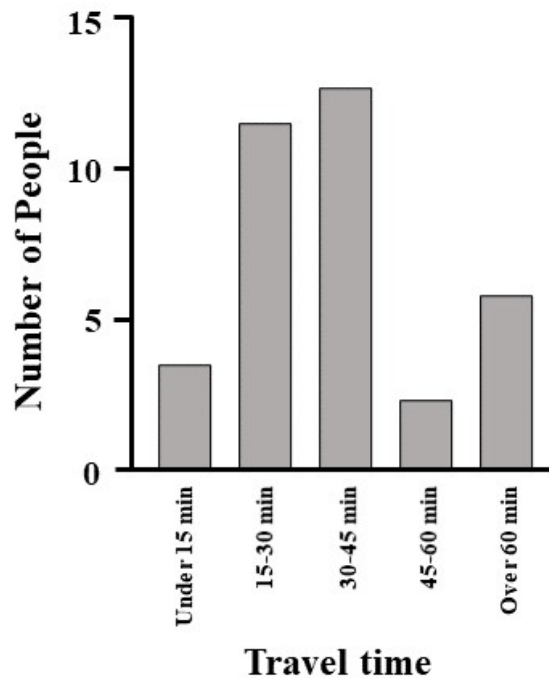
Based on our survey on the implementation of web conference systems for IRB meetings, we were able to clarify certain aspects related to such meetings. For instance, there was no significant change in facilities that delayed IRB meetings, even in facilities located in a COVID-19 endemic area. There was also no significant increase in the rates of implementing web conference systems for IRB meetings at facilities located in a COVID-19 endemic area. Furthermore, the introduction rate of web conference systems was high in facilities that digitized their review materials (Figure 3). The digitization of review materials may have influenced the introduction of web conference systems for IRB meetings. In addition, 71% (15/21) of the facilities spent US $0 in the introduction of such meetings (Table 1). This indicated that facilities that did not adopt a web conference system would have incurred almost no financial cost in implementing a web conference system. This may be because

90% (19/21) of the facilities held meetings using personal devices. In addition, internet connectivity was identified as the most common issue in the introduction of web conference systems. However, this issue can be resolved in the future with advancements in science and technology [7,8]. Our results suggest that financial cost is not an obstacle in the introduction of web conference systems for IRB meetings. However, since such meetings rely on an internet connection, it is necessary to carry out measures for unanticipated issues that are nonexistent during face-to-face meetings.

In our survey, the web conference systems varied depending on the facility, and even the most used system, WebEX, was only used by 38% (8/21) of facilities. Further, if IRB information is compromised, many facilities, such as companies and universities, would suffer great losses. Security measures are important for preventing information leakage, but these

measures are not standardized among facilities (Table 2). Nevertheless, all security measures implemented by the surveyed facilities were considered appropriate. In order to prevent information leakage in facilities with poor security measures, the country's regulatory boards and organizations are expected to release guidelines concerning the improvement of implementing web conference systems, including guidelines for security measures.

Of the 21 facilities that introduced web conference systems for IRB meetings, 15 (71%) stated that they would continue to hold web-based meetings. Therefore, it is likely that such IRB meetings will continue, even after the pandemic, when assembly IRB meetings can be resumed. However, there are concerns that the use of conferencing systems will reduce the quality of IRB meeting deliberation. Our survey was conducted to clarify the measures needed to improve the quality of IRB meetings conducted through web conference systems.

This study revealed that the use of cameras during meetings, the content displayed on computer screens, and the scope of members who participate in meetings using the web conference system differed depending on the facility (Table 3). Additionally, compared to in-person IRB meetings conducted before the COVID-19 pandemic, no facility shortened the time of discussion by more than 30 minutes for web-based IRB meetings, and 86% of facilities stated that IRB members made the same number of comments during such meetings (Table 4). Based on these results, in order to improve the quality of IRB meetings conducted through web conference systems, it is necessary to further examine members' use of cameras, the content displayed on members' computer screens, and the scope of members who participate in meetings using a web conference system.

Due to the COVID-19 pandemic, the overall number of reviews in clinical trial core hospitals has decreased. Additionally, while 68% (21/31) of surveyed facilities introduced web conference systems for IRB meetings, 13% (4/31) postponed IRB meetings. According to the 71% (15/21) of the respondents who implemented web conference systems for IRB meetings, the cost of implementation was US $0, showing that there is almost no financial burden for implementing web conference systems. In 90% (19/21) of the facilities, web-based IRB meetings were already being conducted using personal electronic devices.

Furthermore, in 48% (10/21) of the facilities, the web conference system was used in conjunction with face-to-face meetings. This study revealed that the use of cameras during meetings, the content displayed on computer screens, and the scope of members who participate in meetings using a web conference system differed depending on the facility.

Based on our results, we found that although many facilities could create a system for conducting web-based IRB meetings, the methods for implementing such meetings were not uniform among facilities. It is necessary to ensure the quality of IRB discussions, even if a web conferencing system is used. Further studies on the use of cameras, the content displayed on computer screens, and the scope of members who participate in meetings using a web conference system are necessary to ensure the quality of the system.

Limitations

This study may have some limitations. For instance, our survey was administered to hospitals in Japan. Since the laws regarding clinical trials vary from country to country, the measures obtained in this study may not necessarily apply to other countries. In addition, since the survey was conducted under special circumstances (ie, the COVID-19 pandemic), it is necessary to carefully discuss whether our findings can be applied to situations after the COVID-19 pandemic. Some respondents stated that the attendance rate increased due to the introduction of web conference systems for IRB meetings, but it cannot be denied that this may be due to the decrease in the number of business trips and other meetings caused by the impact of COVID-19. As such, the attendance rate during web-based IRB meetings must be reviewed after the end of the pandemic.

Conclusions

In this study, we identified the problems in conducting IRB meetings through web conference systems. It is necessary to improve the quality of such IRB meetings by investigating and verifying the measures for solving these problems. The results of this study can be used to guide future IRB meetings held after the end of the COVID-19 pandemic, once it becomes possible to hold face-to-face meetings. Our results are not limited to IRBs and can be used as a reference for introducing a web conference system to other organizations.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Table S1. The questions in the Questionnaire.
[DOCX File , 19 KB - jmir_v22i11e22302_app1.docx]

References

1. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, WHO Strategic and Technical Advisory Group for Infectious Hazards. COVID-19: towards controlling of a pandemic. Lancet 2020 Mar 28;395(10229):1015-1018 [FREE Full text] [doi: 10.1016/S0140-6736(20)30673-5] [Medline: 32197103]

2. Ong SWX, Tan YK, Chia PY, Lee TH, Ng OT, Wong MSY, et al. Air, Surface Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient. JAMA 2020 Apr 28;323(16):1610-1612 [FREE Full text] [doi: 10.1001/jama.2020.3227] [Medline: 32129805]

3. Declaration of the State of Emergency Over the Novel Coronavirus. Head of the Novel Coronavirus Response Headquarters. URL: https://jhep.jp/jhep/sisetu/pdf/tokuken2020_15.pdf [accessed 2020-11-09]

4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline, Guideline for Good Clinical Practice E6(R1). URL: https://www.pmda.go.jp/files/000156725.pdf [accessed 2020-11-09]

5. Pharmaceuticals and Medical Devices Agency. Ministerial Ordinances. URL: https://www.pmda.go.jp/english/review-services/regulatory-info/0001.html [accessed 2020-11-09]

6. On handling of reviews by the Institutional Review Board regarding clinical trials related to the novel coronavirus. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/hourei/doc/tsuchi/T200403I0040.pdf [accessed 2020-11-09]

7. ITU. Yearbook of Statistics, Telecommunication/ICT Indicators 2009-2018. Geneva: ITU Publications; 2020.

8. Broadband Portal. Organisation for Economic Co-operation and Development. 2020 Jul. URL: http://www.oecd.org/sti/broadband/broadband-statistics/ [accessed 2020-11-09]

9. Porpiglia F, Checcucci E, Autorino R, Amparore D, Cooperberg MR, Ficarra V, et al. Traditional and Virtual Congress Meetings During the COVID-19 Pandemic and the Post-COVID-19 Era: Is it Time to Change the Paradigm? Eur Urol 2020 Sep;78(3):301-303 [FREE Full text] [doi: 10.1016/j.eururo.2020.04.018] [Medline: 32334884]

10. Gates B. Responding to Covid-19 - A Once-in-a-Century Pandemic? N Engl J Med 2020 Apr 30;382(18):1677-1679. [doi: 10.1056/NEJMp2003762] [Medline: 32109012]

11. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]

12. Gunasekeran DV, Liu Z, Tan WJ, Koh J, Cheong CP, Tan LH, et al. Evaluating Safety and Efficacy of Follow-up for Patients With Abdominal Pain Using Video Consultation (SAVED Study): Randomized Controlled Trial. J Med Internet Res 2020 Jun 15;22(6):e17417 [FREE Full text] [doi: 10.2196/17417] [Medline: 32459637]

13. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]

14. Salisbury C, Murphy M, Duncan P. The Impact of Digital-First Consultations on Workload in General Practice: Modeling Study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]

15. Vandekerckhove P, Vandekerckhove Y, Tavernier R, De Jaegher K, de Mul M. Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights. J Med Internet Res 2020 Jun 25;22(6):e19771 [FREE Full text] [doi: 10.2196/19771] [Medline: 32519964]

16. Kies JK, Williges RC, Rosson MB. Evaluating desktop video conferencing for distance learning. Comput Educ 1997 Feb;28(2):79-91. [doi: 10.1016/s0360-1315(97)00004-3]

17. Angeli C, Valanides N, Bonk C. Communication in a web-based conferencing system: The quality of computer-mediated interactions. Br J Educ Technol 2003 Jan;34(1):31-43. [doi: 10.1111/1467-8535.00302]

18. Reushle S, Loch B. Conducting a trial of web conferencing software: Why, how, and perceptions from the coalface. Turkish Online Journal of Distance Education 2008 Jul;9(3):19-28 [FREE Full text]

19. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

20. Fukushima M, Austin C, Sato N, Maruyama T, Navarro E, Rocca M, et al. The Global academic research organization network: Data sharing to cure diseases and enable learning health systems. Learn Health Syst 2018 Dec;3(1):e10073 [FREE Full text] [doi: 10.1002/lrh2.10073] [Medline: 31245596]

Abbreviations

IRB: institutional review board

XSL•FO
RenderX

Original Paper

A SARS-CoV-2 Surveillance System in Sub-Saharan Africa: Modeling Study for Persistence and Transmission to Inform Policy

Lori Ann Post[1], PhD; Salem T Argaw[2], BA; Cameron Jones[3], MPH, MD; Charles B Moss[4], PhD; Danielle Resnick[5], PhD; Lauren Nadya Singh[1], MPH; Robert Leo Murphy[6], MD; Chad J Achenbach[3], MD; Janine White[1], MSc; Tariq Ziad Issa[2], BA; Michael J Boctor[2], BSc; James Francis Oehmke[1], PhD

[1]Buehler Center for Health Policy & Economics and Departments of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Division of Infectious Disease, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[4]Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

[5]International Food Policy Research Institute, Washington, DC, United States

[6]Institute for Global Health, Northwestern University, Chicago, IL, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy & Economics and Departments of Emergency Medicine
Feinberg School of Medicine
Northwestern University
750 N Lake Shore Dr
9th Floor, Suite 9-9035 Rubloff Building
Chicago, IL, 60611
United States
Phone: 1 203 980 7107
Email: lori.post@northwestern.edu

Abstract

Background: Since the novel coronavirus emerged in late 2019, the scientific and public health community around the world have sought to better understand, surveil, treat, and prevent the disease, COVID-19. In sub-Saharan Africa (SSA), many countries responded aggressively and decisively with lockdown measures and border closures. Such actions may have helped prevent large outbreaks throughout much of the region, though there is substantial variation in caseloads and mortality between nations. Additionally, the health system infrastructure remains a concern throughout much of SSA, and the lockdown measures threaten to increase poverty and food insecurity for the subcontinent's poorest residents. The lack of sufficient testing, asymptomatic infections, and poor reporting practices in many countries limit our understanding of the virus's impact, creating a need for better and more accurate surveillance metrics that account for underreporting and data contamination.

Objective: The goal of this study is to improve infectious disease surveillance by complementing standardized metrics with new and decomposable surveillance metrics of COVID-19 that overcome data limitations and contamination inherent in public health surveillance systems. In addition to prevalence of observed daily and cumulative testing, testing positivity rates, morbidity, and mortality, we derived COVID-19 transmission in terms of speed, acceleration or deceleration, change in acceleration or deceleration (jerk), and 7-day transmission rate persistence, which explains where and how rapidly COVID-19 is transmitting and quantifies shifts in the rate of acceleration or deceleration to inform policies to mitigate and prevent COVID-19 and food insecurity in SSA.

Methods: We extracted 60 days of COVID-19 data from public health registries and employed an empirical difference equation to measure daily case numbers in 47 sub-Saharan countries as a function of the prior number of cases, the level of testing, and weekly shift variables based on a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: Kenya, Ghana, Nigeria, Ethiopia, and South Africa have the most observed cases of COVID-19, and the Seychelles, Eritrea, Mauritius, Comoros, and Burundi have the fewest. In contrast, the *speed, acceleration, jerk, and 7-day persistence* indicate rates of COVID-19 transmissions differ from observed cases. In September 2020, Cape Verde, Namibia, Eswatini, and South

Africa had the highest speed of COVID-19 transmissions at 13.1, 7.1, 3.6, and 3 infections per 100,0000, respectively; Zimbabwe had an acceleration rate of transmission, while Zambia had the largest rate of deceleration this week compared to last week, referred to as a *jerk*. Finally, the 7-day persistence rate indicates the number of cases on September 15, 2020, which are a function of new infections from September 8, 2020, decreased in South Africa from 216.7 to 173.2 and Ethiopia from 136.7 to 106.3 per 100,000. The statistical approach was validated based on the regression results; they determined recent changes in the pattern of infection, and during the weeks of September 1-8 and September 9-15, there were substantial country differences in the evolution of the SSA pandemic. This change represents a decrease in the transmission model R value for that week and is consistent with a de-escalation in the pandemic for the sub-Saharan African continent in general.

Conclusions: Standard surveillance metrics such as daily observed new COVID-19 cases or deaths are necessary but insufficient to mitigate and prevent COVID-19 transmission. Public health leaders also need to know where COVID-19 transmission rates are accelerating or decelerating, whether those rates increase or decrease over short time frames because the pandemic can quickly escalate, and how many cases today are a function of new infections 7 days ago. Even though SSA is home to some of the poorest countries in the world, development and population size are not necessarily predictive of COVID-19 transmission, meaning higher income countries like the United States can learn from African countries on how best to implement mitigation and prevention efforts.

Introduction

Background

In December 2019, a novel illness leading to severe pneumonia and acute respiratory disease was observed in Wuhan, China. The etiologic pathogen was identified as a novel coronavirus, SARS-CoV-2, and the associated respiratory illness was named COVID-19 [1]. As cases began to spread from China, the outbreak was recognized as a global health emergency by the World Health Organization (WHO). Within weeks, the economic significance, clinical manifestations, and the burden on local health systems began to be felt universally. To date (September 27, 2020), over 33 million cases have been reported worldwide, with 128.1 deaths per 1 million [2]. Africa as a whole has reported 1.45 million cases, while the United States has reported over 7.2 million cases, India over 5.9 million, and Brazil over 4.7 million [2].

Using commonalities in geography, development, economies, and climate, the World Bank encompasses 48 countries within the sub-Saharan Africa (SSA) region [3]. The countries of SSA today (September 27, 2020) report 1,131,385 total COVID-19 cases and 26,285 COVID-19–related deaths [2]. SSA comprises immense diversity among the 750 million people that reside in the region. Africa has more than 2000 separate and distinct languages, as well as numerous cultural and ethnic groups with their own history, traditions, and religion [4]. These countries also vary in governance, political stability and unrest, infrastructure, medical resources, and public health [5]. Given their proximity and heterogeneity, SSA countries face additional challenges implementing unified policies to combat the pandemic.

The first reported case of COVID-19 in SSA occurred in Nigeria in February 2020 (Figure 1) [6]. The case was an Italian citizen who returned to Lagos, Nigeria for work after visiting Italy. The Nigerian Minister of Health published health precautions following the positive test result encouraging all Nigerians to take safety measures such as hand washing and social distancing seriously [6]. Since then, Nigeria has reported over 55,000 cases with 5 deaths per 1 million [2]. South Africa has thus far reported the greatest number of cases in SSA at nearly 642,000 cases, with countries like Seychelles, Eritrea, and Mauritius reporting the least for the region at less than 500 cases [2].

SSA has vast experience in tracking and treating infectious outbreaks, with 45% of the continent facing at least one epidemic per year [7]. Countries have been able to leverage this experience to effectively ramp up diagnostic testing for COVID-19 and reinforce previously communicated public health practices such as hand hygiene and social distancing [8]. In Nigeria, tuberculosis (TB) and HIV testing technologies have been converted for COVID-19 use, and in Ethiopia, Abbott reconfigured instruments to increase COVID-19 testing capacity to 7600 tests per day [9]. Through such measures, African countries increased diagnostic capacity from 2 to 43 countries in 3 months and are currently working to further increase subnational capacity [9].

Urban areas must also mitigate the transmission of the virus within a context of high population density and mobility. The

rapid urbanization of SSA cities has resulted in nearly 60% of urban dwellers (336 million people) living in informal settlements and slums [10]. SSA must cope with all of these challenges along with the burden of its inconsistent infrastructure and policy systems (ie, inequalities in access to safe water and sanitation, drought impacts, access to health care facilities), high population to medical provider ratio, and fragile economies vulnerable to lockdown costs in facing this pandemic [10].

Figure 1. Sub-Saharan Africa COVID timeline.
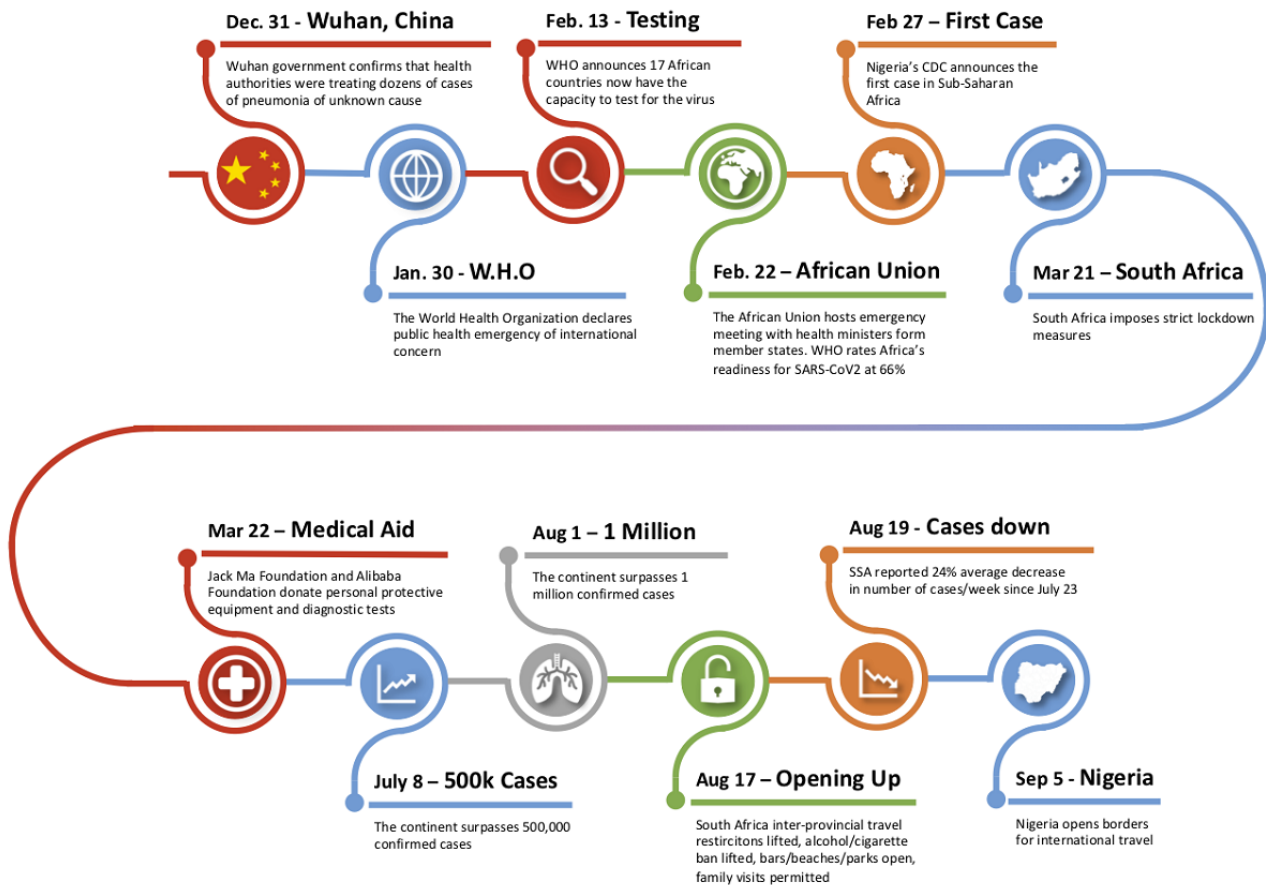
Demographic and Geographic Profile

SSA has high fertility combined with a decrease in infant and child mortality, which results in a short and wide base in the population pyramid (see Figure 2) [11]. Compared to high income and developed countries, SSA is a youthful population. SSA is home to 37 of the world's youngest countries with nearly 60% of the population younger than 25 years and more than 90% of its population younger than 50 years [12] (see Figure 2). Although younger people are less likely than older adults to develop severe outcomes of COVID-19, it is unclear if SSA's youthful population benefits from age-related protective factors due to the health effects of food insecurities and the high communicable disease rate (eg, HIV/AIDS, TB, and malaria) [12]. Some research suggests that the young average age of the population and the prominence of rural dwellings may limit the overall spread and severity of the pandemic, perhaps outweighing the negative impact incurred by the high communicable disease burden [5]. The pervasiveness of rural dwellers, however, also increases the likelihood of staggered outbreaks, requiring vigilance in policy making [5]. Children are just as likely as adults to become infected with SARS-CoV-2 but are less likely to be symptomatic or develop severe symptoms [13-15]. Thus, when comparing SSA to Europe, Asia, and North America, SSA has a significantly younger population structure [16].

Population density is another COVID-19 demographic risk factor. The closer people live to each other and the more they interact, the greater the likelihood of transmitting the novel coronavirus [17]. Even though the African continent has experienced significant growth due to the high fertility rates and lower mortality rates, Africa still has several sparsely populated areas. In 2017, the average population density was 16.6 persons per square kilometer. Today, the average population density of the whole world is 51 persons per square kilometer, and 23 African countries rank below the global average. Moreover, most of the African countries with the highest population density are smaller countries such as Rwanda, Comoros, Burundi, and Mauritius [18].

Out of the more than 1 billion people that live in SSA, there are 2,213,327 more females than males. Worldwide, males generally have experienced worse outcomes from COVID-19 than females [19-21].

Figure 2. Population pyramid for sub-Saharan Africa.

Age	Male	Female
100+	0.0%	0.0%
95-99	0.0%	0.0%
90-94	0.0%	0.0%
85-89	0.0%	0.1%
80-84	0.1%	0.2%
75-79	0.3%	0.3%
70-74	0.4%	0.5%
65-69	0.7%	0.8%
60-64	0.9%	1.0%
55-59	1.2%	1.3%
50-54	1.5%	1.6%
45-49	1.9%	2.0%
40-44	2.3%	2.4%
35-39	2.9%	2.9%
30-34	3.4%	3.4%
25-29	3.9%	3.9%
20-24	4.5%	4.4%
15-19	5.2%	5.1%
10-14	6.0%	5.9%
5-9	6.9%	6.7%
0-4	7.7%	7.4%

PopulationPyramid.net

AFRICA - 2019
Population: **1,308,064,176**
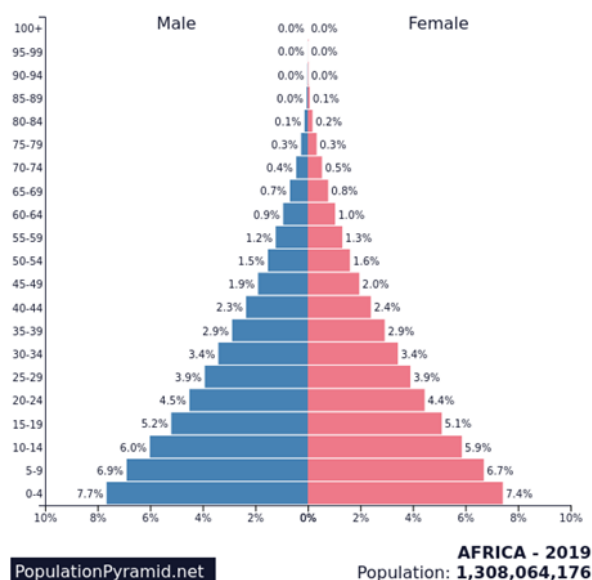
Policies and Governance

As early as January 2020, countries such as Côte d'Ivoire (Ivory Coast) began implementing symptom checks at airports and travel restrictions from China [22]. In February, the Africa Centres for Disease Control and Prevention (Africa CDC) organized the *Africa Joint Continental Strategy: COVID-19 Outbreak among the African Union*. Member states created an international task force aimed at preventing mortality and morbidity while abating social disturbance and economic penalties [23]. This task force has since recommended evidence-based policies, worked to procure diagnostic tools, organized workshops for containment and mitigation strategies, and provided training and advice for public health systems [8,22]. By the end of May, most African countries had banned flights from Asia and Europe as well as implemented quarantine periods for other travelers and placed restrictions on school and other public gatherings, in addition to executing lockdowns and curfews [22].

Although most SSA countries acted promptly and decisively, others have been more inert. Tanzania's President John Magufuli, for example, declared diagnostic tests defective in May after samples from goats and sheep returned positive [24]. Soon after, the country abandoned reporting new cases (last reported number at 509) and in June declared itself virus-free [25]. Subsequently, neighboring countries (ie, Kenya, Uganda, and Zambia) tightened their borders in an attempt to control spread, especially with truck driver mobility [24]. In addition, Uganda and Kenya set curfews and partial or full lockdowns alongside designated isolation facilities [24,26]. Similarly, Zambia set aside two medical facilities in the capital for the purpose of quarantining, and the Democratic Republic of the Congo closed provincial borders, churches, and educational facilities in addition to implementing curfews [26,27]. This was in contrast to Tanzania's open churches and mosques as well as public transportation despite an initial ban on public gatherings [28,29]. Still, in other countries the spread of false or unconfirmed information by politicians and public figures about coronavirus "cures" such as the herbal tonic promoted by President Andry Rajoelina of Madagascar caused confusion and some preventable deaths [30].

Another major concern has been the use of oppressive actions by law enforcement to contain outbreaks. There have been reports in Kenya of civilian deaths by security forces in the process of enforcing curfews [31]. Uganda similarly reported attacks on lesbian, gay, bisexual, transgender, and queer individuals in a shelter for failure to social distance [31]. Uganda, Nigeria, and the Democratic Republic of Congo have seen the largest increase in violence targeting civilians since the pandemic onset compared to the immediate prepandemic period [32]. At the expense of civil liberties, South Africa and Mozambique set restrictions on the spread of misinformation around the outbreak. Ethiopia postponed elections in an effort to minimize large gatherings, and Ghana banned political rallies in advance of its December 2020 elections [12,33]. Achieving the delicate balance between protecting human rights and limiting misinformation while effectively containing outbreaks in countries with a history of oppressive governments and political unrest will be critical to regaining stability in the postcoronavirus era.

Health Systems

Public health systems must confront provider shortages, limited health facilities, and medical supply shortages in meeting community needs [34]. Rural residents especially have to travel longer distances and have worse access to health care due to these shortages [34]. On average, African countries have 1.8 hospital beds per 1000 people. South Africa has 2.8, Botswana 1.8, Nigeria 0.5, Ethiopia 0.3, and Ghana 0.9 [35]. Uganda has 55 intensive care unit beds for 40 million people with added limitations due to a dearth in critical care doctors and anesthetists [36]. South Africa has 0.9 physicians per 1000 people, Botswana 0.5, Nigeria 0.4, and Ethiopia and Ghana 0.1 [35]. In comparison, the United States has 2.6, Italy has 4.0, and China has 2.0 physicians per 1000 people [37]. Armed conflicts in South Sudan after nearly a decade of conflict have

weakened the health system with damage to infrastructure, depletion of workforce, and reallocation of funds toward national security [38]. In Burkina Faso, displacement of people and increased migration across its borders has similarly complicated the health system's ability to meet health needs [38].

Furthermore, many SSA countries are not able to screen for COVID-19 using constitutional symptoms like fevers, as they face high concurrent infection rates [5,36]. The current pandemic necessitates redirection of resources from other public health initiatives to COVID-19, allowing countries to meet imminent needs at the expense of others. Countries like Nigeria have been able to effectively combat COVID-19 by transitioning infrastructure (ie, laboratory services, surveillance, and human resources) and converting diagnostic tests previously meant for poliovirus, as they had done during the Ebola outbreak [8,38,39]. Although this is innovative and highly effective, it does come with consequences: the Ebola outbreak was later associated with a decrease in hospital births and an increase in maternal mortality rates [10]. Recent news has similarly revealed major concerns around maternal mortality in light of COVID-19 precautions [40].

There is a global shortage of personal protective equipment (ie, face masks, plastic and rubber gloves, and surgical garments), diagnostic tools, and ventilators, which is more pronounced in SSA due to dependencies on imported medical and pharmaceutical products [38]. African countries as a whole import a total of US $748 million in protective garments and US $9291 million in disinfectants and sterilization products annually [10]. Countries worldwide have placed limitations on the export of supplies they deem necessary for tackling COVID-19 internally. These include protective masks, ventilators, medicines, disinfectants and sterilization products, intubation products, and paper bedsheets among others [10]. Although this adds to the concerns for the region, the WHO, Jack Ma Foundation, and Alibaba Foundation in Ethiopia, among many others, have continued to donate medical equipment and diagnostic kits to mitigate burdens on African Union states [26,41].

Economics

The World Bank estimates 70-100 million people worldwide may be pushed below the poverty level as a result of the COVID-19 pandemic, and anywhere from 26-39 million of them reside in SSA, which is now projected to experience its first recession in 25 years [42]. The economic effects of the pandemic are likely to impact younger people in the sub-Saharan workforce who provide physical labor [43]. In South Africa, the government has imposed several waves of lockdowns and closures, and last month was forced to reimpose a curfew when cases rebounded. The South African Chamber of Commerce and Industry has projected unemployment as high as 50% in coming months, and the economy is projected to shrink 7% this year [44]. The novel coronavirus has also delayed the implementation of the African Continental Free Trade Area implementation and instead has led to a series of border closures throughout the region [45]. The economic policy responses to COVID-19 have varied significantly between nations and have involved a number of different tactics: interest rate cuts, fee

waivers, and direct relief. A majority of sub-Saharan national governments are now offering some level of income support for lost wages as a result of the pandemic, though only Gabon and Benin are providing more than 50% of lost salary. In addition, a number of the region's most populous nations including Nigeria, Ethiopia, the Democratic Republic of Congo, Tanzania, and South Africa, are currently offering no wage relief [46].

Food Security

The impact of COVID-19 on food security is both direct and indirect: agricultural production itself is projected to contract from 2% to 7% alongside decreased food imports resulting from increased costs associated with border closures and disrupted supply chains. An analysis of Ghana's 3-week lockdown in March found the agrifood system gross domestic policy, which incorporates primary agriculture as well as food trade and food service industries, declined 19.8%, even though primary agricultural activities were excluded from lockdown measures [47]. Nigeria also saw a similar drop in agrifood gross domestic product during its own lockdown [48]. The primary reason for a projected increase in acute hunger is simply falling incomes resulting from the economic fallout previously described [42]. Those near or below the poverty line spend a larger share of their income on food, and any decline in incomes is likely to directly impact their access to food [49]. Food prices have begun to rise, which may leave struggling residents with fewer calories and a need to substitute with cheaper, less nutritious options. Indeed, a study conducted in Addis Ababa, Ethiopia in May found that one-fourth of all households reported reducing their household food expenditures over the previous month, and the biggest decline in spending was on fresh produce [50]. Eastern Africa countries including Ethiopia, Kenya, and Tanzania have faced large-scale crop devastation due to locusts and flooding. Budget shortfalls have limited food distribution by the World Food Programme and the United Nations Commissioner for Refugees in the region, affecting the 3.2 million refugees currently residing in Ethiopia, Kenya, Sudan, South Sudan, and Tanzania. A recent survey of South Sudanese refugees found 80% were resorting to rationing or skipping meals, and refugee camps in Rwanda have reported a 27% increase in food prices. So-called transitional food supply chains, which feature many links and often end at small enterprises and wet markets, are the predominant pattern in much of SSA and will likely struggle to comply with stringent government-imposed regulation [51].

Sociocultural

There are a number of social and cultural factors that affect both the impact of COVID-19 on SSA and the response to the pandemic. Infrastructure challenges pose a barrier on a number of fronts: a lack of piped water limits hand hygiene, uneven personal transportation options impede access to resources during lockdowns, and overcrowded housing in urban and semiurban areas make social distancing impossible [7]. There is also a fear that lockdowns and closures may exacerbate certain structural inequalities. School closures threaten long-term educational attainment, and socially distant learning is simply not an option in locales with limited or nonexistent internet access. According to data from the International

Telecommunications Union, Africa has the largest share of the world's offline population, with an overall internet penetration rate of around 28% [52]. Gender-based violence is also expected to increase with prolonged lockdowns, and the pandemic threatens to reverse hard-fought economic and social advances in gender equality in recent decades [40,53,54]. A final psychosocial factor playing out in real time is the impact of recent previous infectious disease outbreaks such as Ebola and HIV. Although this recent experience has in some cases led to more aggressive policy responses than in other global regions, many of the public health response challenges for Ebola, as well as the mistrust of health care workers and stigmatization of victims and survivors, have persisted amid the COVID-19 response [55].

Long-Term Outcomes of COVID-19

The legacy of the SARS-CoV-2 pandemic in SSA will depend on the morbidity and mortality of the pandemic but also the response. Currently, South Africa accounts for over half of reported cases in SSA, and although limited testing and surveillance may put those numbers into question, it appears that much of the region has yet to experience a large outbreak [56,57]. As previously outlined, many of the region's countries acted decisively and swiftly to impose lockdown measures even as case numbers were still relatively low. The region now, however, faces the ongoing dilemma of confronting worsening poverty and food insecurity while also maintaining some public health measures to limit a renewed wave of infection. The lack of health care infrastructure remains a significant concern, and the informal and labor-intensive character of the region's economy makes lockdowns, that wealthier nations implement, unsustainable. Public health experts fear the upshot of the pandemic may be a resurgence in malaria, TB, and other deadly diseases [58]; however, there remains hope that the response to COVID-19 and the upheaval resulting from the virus may facilitate creative and inclusive initiatives that not only limit the effects of the virus but also contribute to a sustainable recovery and growth [12].

Scientists continue to praise Africa's quick response to closing international borders and enforcing strict quarantine and social distancing measures in many countries even before any deaths were reported [59,60]. Since August 1, 2020, nations in SSA have seen an attenuation in many of the public health and social measures responsible for the incredible early success in combatting the virus [61]. Countries have begun to lift their lockdowns and ease back into a prepandemic form of life. Earlier in September, Zimbabwe recorded no COVID-19–related deaths over an entire week [61]; moreover, many residents have stopped wearing masks, and the government eased its restrictions. As Nigeria similarly began to see a decline in cases, they opened their borders for international travel on September 5, 2020 [62].

With reported cases falling in South Africa, the government reduced its national alert level from level 3 to level 2 on August 17, 2020, lifting travel restrictions between provinces; allowing bars, taverns, gyms, and beaches to open; and ending the ban of alcohol and cigarette sales [63,64]. After continued reduced disease spread, the government further reduced its national alert level to level 1 on September 20, 2020, increasing the capacity allowed within event venues [65]. As these restrictions continue to be eased across the continent, their effectiveness and impact on the spread of SARS-CoV-2 are still to be determined.

Objective

Several organizations have developed surveillance and tracking tools for SARS-CoV-2 cases in Africa. These include the WHO Regional Office of Africa, the Africa CDC, the Worldometer, the Tony Blair Institute for Global Change, Johns Hopkins University, Statista, and the New York Times [2,7,22,66-69]. Although these data repositories offer good proxies for COVID-19 morbidity and mortality, they have undercounts, errors, and reporting biases. To that end, the objective of our research is to derive speed, acceleration and deceleration, jerk, and 7-day persistence to measure where and how rapidly COVID-19 is transmitting. We will rely on dynamic panel modeling and generalized method of the moments that correct for data limitations. Our study will measure the underlying causal effect from the previous week that persists through the current week with a 7-day persistence rate that explains new infections in the current week as a direct result of infections from the prior week. In SSA, for example, a 7-day persistence can occur following protest against postponed elections as was seen in Guinea this March. In summary, we will measure negative and positive shifts in the transmission of SARS-CoV-2 or the acceleration and deceleration rates that do not have sampling bias.

Methods

The Foundation for Innovative New Diagnostics [70] compiles data from multiple sources across individual websites, statistical reports, and press releases; data for the most recent 8 weeks were accessed from the GitHub repository [71]. This resulted in a panel of 47 countries with 60 days in each panel (n=2820). An empirical difference equation was specified in which the number of positive cases in each state at each day is a function of the prior number of cases, the level of testing, and weekly shift variables that measure whether the contagion was growing faster, slower, or stayed the same than the previous weeks. This resulted in a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R (the R Foundation for Statistical Computing) [72,73].

Arellano-Bond estimation of difference equations has several statistical advantages: it allows for statistical examination of the model's predictive ability and the validity of the model specification, it corrects for autocorrelation and heteroscedasticity, it has good properties for data with a small number of time periods and large number of countries, and it corrects for omitted variables issues and provides a statistical test of correction validity. With these advantages, the method is applicable to ascertaining and statistically validating changes in the evolution of the pandemic within a period of a week or less, such as changes in the reproduction rate [74].

Results

Country Regression Results

We group 47 countries into the global region of SSA and present the regression results. The weekly surveillance products will be based on these regressions.

For SSA, the regression Wald statistic showed that the model was statistically significant (χ^2_4=289, P<.001), and the Sargan test failed to reject the validity of the overidentifying restrictions χ^2_{778}=47, P>.99; Table 1).

Table 1. Arellano-Bond dynamic panel data modeling of the number of daily infections reported by country, September 1-15, 2020.

Variable	Coefficient	P value
L1Pos[a]	0.052	.20
L7Pos[b]	0.069	.11
Cumulative Tests	0.013	<.001
Shift parameter week of September 8	–0.359	<.001
Shift parameter week of September 15	–1.05	<.001
Weekend	–0.131	.19

[a]L1Pos: 1-day persistence. The number of new infections today that are statistically related to those new infections 1 day ago per 100,000 population. Infections 1 day ago has an echo forward effect in COVID-19 transmissions that we call 1-day persistence rate.

[b]L7Pos: 7-day persistence. The number of new infections today that are statistically related to those new infections 7 days ago per 100,000 population. Infections 7 days ago has an echo forward effect in COVID-19 transmissions that we call 7-day persistence rate.

The coefficient on the first persistence of the dependent variable was not statistically significant, nor were the shift parameters for the weeks of September 8 and 15, 2020, for this coefficient. The coefficient on the seventh persistence was not statistically significant. This is consistent with the COVID-19 transmission rates. Overall, the rates are not changing significantly in SSA. The shift parameter for the weeks of September 8 and 15 were statistically significant. The cumulative number of tests administered was significant (coefficient 0.013, P=.001). The weekend variable was significant (coefficient –0.131, P=.19).

Interpretation: SSA Regression Results

SSA overall appears to be fairly calm, with the only statistically significant effect being cumulative testing and the weekend effect. Both static surveillance metrics such as new cases of COVID-19 infections, cumulative cases, and deaths indicate a leveling or declining effect, but some countries appear to have accelerated growth in rates of COVID-19 transmissions.

Surveillance Results

Surveillance results are presented in Tables 2-5. The four novel surveillance indicators are calculated as weekly averages to compare the transmission of COVID-19 from week to week (see Tables 2 and 3). These surveillance system data elements include average weekly number of daily positive tests per 100,000 population, referred to as *speed*; weekly average of day-to-day change in the number of positives per day per 100,000 population, referred to as *acceleration*; change in acceleration, referred to as *jerk,* which is the acceleration in the current week less the acceleration in the prior week (a sustained positive jerk is typically associated with explosive growth); and the *7-day persistence*, which is the number of new cases of COVID-19 reported on the current day per 100,000 population (ie, current speed) that are associated with new cases reported 7 days ago (ie, last week's speed), measures how much an increase in speed in the previous week persists into the current week. Data are presented according to SSA countries. These novel indicators are standardized per 100,000 population to compare changes in rates of COVID-19 transmission between countries.

Table 2. Surveillance metrics for the week of September 1-8, 2020.

Country	Speed: daily positives per 100,000 (weekly average of new daily cases per 100,000), n	Acceleration: day-to-day change in the number of positives per day (weekly average per 100,000)	Jerk: week over week change in acceleration (per 100,000)	7-day persistence effect on speed (new cases per day per 100,000), n
Angola	0.14	–0.01	0.00	6.34
Benin	0.08	0.00	0.00	0.43
Botswana	2.49	–0.56	–1.33	2.30
Burkina Faso	0.07	0.00	0.00	0.45
Burundi	0.03	0.00	0.00	0.21
Cameroon	0.24	–0.01	–0.01	6.20
Cape Verde	11.17	–1.14	–1.04	5.71
Central African Republic	0.07	0.02	0.05	0.28
Chad	0.03	0.00	0.00	0.31
Comoros	0.49	0.00	0.00	0.14
Congo-Brazzaville (Republic of Congo)	2.42	0.00	–0.70	0.00
Congo-Kinshasa (Democratic Republic of Congo)	0.03	0.01	0.01	3.03
Cote d'Ivoire	0.37	0.02	0.03	7.69
Equatorial Guinea	0.21	–0.25	–0.39	0.55
Eritrea	0.09	0.04	0.05	0.06
Eswatini (Swaziland)	3.56	–0.26	–0.45	4.14
Ethiopia	0.95	0.00	0.00	136.70
Gabon	0.49	0.00	0.14	1.76
Gambia	1.50	0.07	0.07	4.88
Ghana	0.26	–0.01	–0.01	10.56
Guinea	0.41	–0.04	–0.02	4.99
Guinea-Bissau	0.30	0.00	0.00	0.80
Kenya	0.28	0.01	0.02	21.50
Lesotho	0.42	0.00	0.00	0.51
Liberia	0.02	0.00	–0.01	0.14
Madagascar	0.25	–0.01	–0.01	6.85
Malawi	0.04	0.00	0.02	2.18
Mali	0.08	0.01	–0.01	0.91
Mauritania	0.28	–0.09	–0.06	1.63
Mauritius	0.06	0.06	0.07	0.11
Mozambique	0.29	–0.02	–0.02	7.55
Namibia	7.08	–0.14	0.21	21.78
Niger	0.00	0.00	0.00	0.04
Nigeria	0.09	0.00	0.00	20.57
Rwanda	0.34	–0.06	–0.05	8.60
Sao Tome and Principe	0.13	0.00	0.00	0.06
Senegal	0.34	–0.01	0.01	8.52
Seychelles	0.15	0.00	–0.15	0.00
Sierra Leone	0.07	0.01	0.00	0.38

Country	*Speed*: daily positives per 100,000 (weekly average of new daily cases per 100,000), n	*Acceleration*: day-to-day change in the number of positives per day (weekly average per 100,000)	*Jerk*: week over week change in *acceleration* (per 100,000)	*7-day persistence effect* on *speed* (new cases per day per 100,000), n
Somalia	0.06	0.01	0.01	0.50
South Africa	2.97	–0.03	0.24	216.68
South Sudan	0.03	0.01	0.02	0.28
Sudan	0.08	0.00	0.00	3.06
Togo	0.17	0.01	0.01	1.52
Tanzania	—[a]	—	—	—
Uganda	0.28	0.02	0.00	8.69
Zambia	0.46	–0.13	–0.12	15.58
Zimbabwe	0.81	0.03	–0.34	5.16

[a]Tanzania does not report COVID-19 cases.

Table 3. Surveillance metrics for the week of September 9-15, 2020.

Country	Speed: daily positives per 100,000 (weekly average of new daily cases per 100,000), n	Acceleration: day-to-day change in the number of positives per day (weekly average per 100,000)	Jerk: week over week change in acceleration (per 100,000)	7-day persistence effect on speed (new cases per day per 100,000), n
Angola	0.24	0.04	0.02	4.32
Benin	0.07	0.00	0.00	0.97
Botswana	2.09	0.00	−0.54	5.71
Burkina Faso	0.19	0.01	0.01	1.36
Burundi	0.01	0.00	0.00	0.30
Cameroon	0.23	−0.11	−0.14	6.24
Cape Verde	13.09	0.60	0.65	6.11
Central African Republic	0.11	−0.02	−0.09	0.34
Chad	0.04	0.00	0.00	0.40
Comoros	0.18	0.10	0.08	0.41
Congo-Brazzaville (Republic of Congo)	0.11	0.00	0.68	12.97
Congo-Kinshasa (Democratic Republic of Congo)	0.02	−0.01	0.00	2.67
Cote d'Ivoire	0.18	−0.02	0.01	9.60
Equatorial Guinea	0.16	0.00	0.09	0.28
Eritrea	0.09	−0.03	−0.03	0.31
Eswatini (Swaziland)	2.79	0.05	0.07	4.07
Ethiopia	0.60	−0.06	0.01	106.34
Gabon	0.30	0.00	−0.03	1.07
Gambia	0.93	−0.33	−0.33	3.50
Ghana	0.30	−0.04	−0.08	7.85
Guinea	0.29	0.02	0.02	5.25
Guinea-Bissau	0.22	0.00	0.00	0.57
Kenya	0.26	−0.01	0.00	14.80
Lesotho	1.20	0.55	0.55	0.90
Liberia	0.05	0.02	0.02	0.09
Madagascar	0.19	−0.03	−0.01	6.80
Malawi	0.05	0.00	0.00	0.77
Mali	0.04	−0.01	0.01	1.49
Mauritania	0.49	0.08	0.09	1.28
Mauritius	0.05	−0.01	−0.01	0.07
Mozambique	0.50	0.07	0.02	8.64
Namibia	5.57	−0.20	−0.05	17.57
Niger	0.00	0.00	0.00	0.03
Nigeria	0.07	−0.01	−0.01	17.19
Rwanda	0.21	−0.01	0.02	4.22
Sao Tome and Principe	0.60	0.07	0.07	0.03
Senegal	0.43	0.17	0.17	5.53
Seychelles	0.44	0.00	0.00	0.01
Sierra Leone	0.11	0.01	0.01	0.51

Country	Speed: daily positives per 100,000 (weekly average of new daily cases per 100,000), n	Acceleration: day-to-day change in the number of positives per day (weekly average per 100,000)	Jerk: week over week change in acceleration (per 100,000)	7-day persistence effect on speed (new cases per day per 100,000), n
Somalia	0.02	–0.01	–0.01	0.87
South Africa	2.70	–0.07	–0.10	173.18
South Sudan	0.05	0.00	–0.01	0.36
Sudan	0.03	0.00	–0.01	3.53
Tanzania	__a	—	—	—
Togo	0.14	–0.01	–0.01	1.38
Uganda	0.39	0.01	–0.02	12.27
Zambia	0.69	–0.01	–0.11	8.12
Zimbabwe	0.18	–0.04	0.40	11.79

[a]Tanzania does not report COVID-19 cases.

Table 4. 7-day persistence difference.

Country	7-day persistence September 8, 2020, n	7-day persistence September 15, 2020, n
Nigera	20.5709689	17.18749
Kenya	21.4950277	14.79916
Namibia	21.7793534	17.57133
Ethiopia	136.703827	106.3378
South Africa	216.684664	173.1828

Table 5. Observed new daily and cumulative COVID-19 infections.

Country	COVID-19 infections September 8, 2020, n	COVID-19 infections September 15, 2020, n
New daily infections		
Kenya	151	__a
Cameroon	244	—
Nigeria	296	—
South Africa	1079	772
Ethiopia	1136	700
Uganda	—	145
Senegal	—	223
Mozambique	—	231
Cumulative COVID-19 infections		
Kenya	35,356	36,301
Ghana	45,012	45,655
Nigeria	55,456	56,478
Ethiopia	60,784	65,486
South Africa	640,441	651,521

[a]Country was not in top 5 for this day.

The innovation of this study is the novel metrics we derived to measure how COVID-19 is spreading and those changes in the rates of transmission [75]. These measures should be considered in combination with traditional static numbers including observed new cases and deaths (see Figure 3). These novel metrics measure how fast the rates are changing [74].

We tracked the rates of COVID-19 transmission for SSA for the weeks of September 1-8 (see Table 2) and September 8-15, 2020 (see Table 3). The president of Tanzania has elected not to report COVID-19 morbidity and mortality, so this study is limited to the remaining 47 countries in SSA. SSA had 4206 new observed cases, with an SSA country average of 90 new cases the week of September 8, 2020, and 3156 new observed cases, with an SSA country average of 67 new observed cases the week of September 15, 2020, with significant variation between countries. The six countries with the highest rates of new infections per 100,000 population the week of September 8, 2020, includes Cape Verde, Namibia, Gambia, South Africa, Eswatini (formerly known in English as Swaziland), and Ethiopia at 7.6, 4.7, 3.3, 1.8, 1.7, and 1.0, respectively (see Table 2); whereas the following week (September 15, 2020), the six countries with the highest rates of new infections per 100,000 population were Cape Verde, Lesotho, Namibia, Eswatini, Senegal, and South Africa at 11.8, 3.9, 3.3, 2.1, 1.4, and 1.3, respectively. Gambia and Ethiopia dropped off the list of the top six countries with highest rates of new infections per 100,000 population, whereas Senegal and Lesotho were added within a 7-day span. Although the number of infections can be a function of time or a superspreader event that results in a one-time increase in COVID-19 infections, it is critical that we measure which country's speed of COVID-19 transmission is *accelerating and decelerating* to inform public health policy. The six countries with the largest acceleration in rates of COVID-19 transmission are Sao Tome, Mauritania, Comoros, Senegal, Lesotho, and Cape Verde. In Table 3, we see that these six countries are accelerating at a rate of 0.066428, 0.075758, 0.100735, 0.169188, 0.551191, and 0.597473 per 100,000 population, respectively (see Table 2). To further identify countries that require immediate COVID-19 transmission mitigation and prevention strategies, we examined the *jerk*. The *jerk* measures the *change in acceleration or deceleration* week over week as the acceleration in the current week minus the acceleration of the previous week. The six highest jerks in positive acceleration rates between the weeks of September 8, 2020, and September 15, 2020, are Equatorial Guinea, Senegal, Zimbabwe, Lesotho, Cape Verde, and Congo-Brazzaville with 0.094818, 0.170064, 0.400904, 0.55119, 0.649427, and 0.682357 per 100,000, respectively (see Table 3). In sum, Cape Verde and South Africa have the highest rate of COVID-19 infections per 100,000 or rate of speed of COVID-19 transmissions per 100,000. Moreover, Senegal, Lesotho, and Cape Verde are rapidly accelerating and have the highest jerk rates, meaning that the acceleration of COVID-19 transmissions has jerked the acceleration rate upwards this past week relative to the previous week.

Controlling for incomplete case ascertainment and data contamination, we derived the 7-day persistence surveillance metric (see Table 4) [75]. The 7-day persistence explains the number of current COVID-19 infections that are a result of the number of COVID-19 infections 7 days ago [75]. Table 4 demonstrates the dynamic nature of the COVID-19 transmissions. On September 8, 2020, the five countries with the highest 7-day persistences were Nigeria, Kenya, Namibia, Ethiopia, and South Africa. For example, on September 8, 2020, South Africa had 216.7 new COVID-19 infections as a result of the number of COVID-19 infections from September 1, 2020. The following week, on September 15, 2020, these same five countries had the highest 7-day persistence based on the underlying condition that resulted in new infections a week ago. Thus, South Africa had 173.1 new cases based on the 1079 cases last week, Ethiopia has 106.3 new COVID-19 infections as a result of the new infections that occurred on September 8, 2020, and so on. The leading five countries with the highest persistence rates have all declined in the current week relative to the previous week.

Rank ordering sub-Saharan African countries by the number of observed daily new infections during the week ending on September 8, 2020, from least to most revealed that Kenya, Cameroon, Nigeria, South Africa, and Ethiopia ranked the highest (see Table 5). Whereas a week later the number of new infections had decreased, and three out of the five countries reporting observed daily infections had displaced Kenya, Cameroon, and Nigeria with Uganda, Senegal, and Mozambique. Another standard static surveillance metric is to enumerate the total number of COVID-19 infections or cumulative infections. The leading five nations with the total highest number of reported cases were Kenya, Ghana, Nigeria, Ethiopia, and South Africa (see Table 5).

The most populous countries in SSA include Kenya, Tanzania, South Africa, the Congo, Ethiopia, and Nigeria (see Table 6). Countries with larger populations are at risk for having more COVID-19 infections by virtue of size. Obviously, Tanzania is not reporting any cases of COVID-19 and, thus, was not part of this study; however, with the exception of the Congo, the other most populous countries in SSA had the most cases of COVID-19 infections over time (see Tables 5 and 6).

For comprehensive surveillance of static or traditional surveillance metrics with novel surveillance metrics for SSA, see Multimedia Appendix 1.

Figure 3. Sub-Saharan Africa region as defined by the United Nations.

Table 6. Most populous sub-Saharan African countries.

Country	Population as of 2020, n
Kenya	52,573,973
Tanzania	58,005,463
South Africa	58,558,270
Congo	86,790,567
Ethiopia	112,078,730
Nigeria	200,963,599

Discussion

Principal Findings

The COVID-19 pandemic poses a threat to public health, economies, and food security around the world. Unfortunately, COVID-19 has reversed some positive gains made in SSA over the past few years in reversing extreme poverty and reducing food insecurity. The best hope for preventing further COVID-19 transmissions and mitigating the current adverse outcomes is through good public health surveillance coupled with enforced policies. Standard surveillance metrics such as daily observed new COVID-19 cases or deaths are necessary but insufficient to mitigate and prevent COVID-19 transmission. Public health leaders also need to know where COVID-19 transmission rates are accelerating or decelerating, whether those rates increase or decrease over short time frames because the pandemic can quickly escalate, and how many current cases are a function of new infections from 7 days ago.

Existing surveillance is helpful as it provides a single lens of the pandemic. For example, Nigeria has the largest population in SSA; South Africa has less than one-third of Nigeria's population (see Table 6), yet Nigeria only has 56,478 cases (~12%) of the number of observed COVID-19 cases that South Africa has reported (n=651,521). These numbers indicate there is some underlying process that has resulted in South Africa having a disproportionate number of COVID-19 cases relative to its population size.

What standard surveillance does not reveal is if the rate of COVID-19 transmission is accelerating or decelerating, or if the rate of acceleration is increasing (jerk), resulting in explosive growth in COVID-19 transmissions. Although the basic reproduction number is a metric of transmission, it is based on missing data, asymptomatic infections, and data contamination [74]. Our analysis of the 7-day persistence that measures the persistence of COVID-19 given the number of new COVID-19 infections a week ago, revealed that the number of COVID-19 infections was declining. Moreover, when we examined the acceleration rate of COVID-19 transmissions in South Africa, we found that the rate of transmission was decelerating. Moreover, during the week of September 15, 2020, there was further evidence that the rate had reversed because the jerk was now negative. The acceleration of the current week, which was negative, minus the rate of acceleration of the previous week was also negative and means the COVID-19 transmission was decelerating and that the rate had increased further. From a public health perspective, even though South Africa has the highest number of total COVID-19 infections since January 2020, the rates of transmission have reversed. There remains a significant and large population of infected persons in South Africa that could result in additional outbreaks without proper adherence to public health guidelines that prevent the spread of COVID-19, but if public health leaders and the public remain resolute with maintaining recommended guidelines and prevention efforts, the numbers of COVID-19 infections will continue to decline.

What standard public health surveillance indicators overlook are those countries who have lower numbers of infections and, thus, do not appear concerning. However, our measures of COVID-19 transmission indicate an alarming acceleration rate in some countries. The six countries with the largest COVID-19 transmission acceleration rates are Sao Tome, Mauritania, Comoros, Senegal, Lesotho, and Cape Verde. Half of these countries are islands, suggesting that initially these island nations were successful at locking down the borders, but now, people with COVID-19 have penetrated the natural barriers that come from being on an island and the novel coronavirus is spreading quickly. Mauritania and Senegal are located in West Africa and share a border, which again suggests that the virus transmission is accelerating after maintaining initial control.

The six highest jerks in positive acceleration rates between the weeks of September 8, 2020, and September 15, 2020, are Equatorial Guinea, Senegal, Zimbabwe, Lesotho, Cape Verde, and Congo-Brazzaville. These six countries lead SSA in increased rates of acceleration this week minus the acceleration from last week, which is indicative of a much quicker rate of transmission that results in explosive growth of COVID-19 transmission rates. From a policy perspective, these countries should increase and enforce strict prevention measures such as quarantine or stay at home, avoiding crowds or large gatherings, social distancing, hand hygiene, and wearing face masks to flatten the curve or reverse current transmission trends. In summary, the number of COVID-19 infections and deaths are important surveillance indicators; however, measuring acceleration and deceleration rates, the jerk in rates, and the 7-day persistence can inform public health policy to mitigate outbreaks and prevent further COVID-19 transmission. Our novel indicators provide additional lens to understand how the pandemic is spreading and to inform policy and actions.

Finally, although we have identified countries that would benefit from informed policies, it merits reckoning that even though SSA is home to some of the poorest countries in the world, development and population size are not necessarily related to the rate and extent of COVID-19 transmission. The youthful population, hot climate, and the number of Africans living and working in rural areas serves as potential protective factors. Of note, this study also confirms the power of public health surveillance and policy. Existing surveillance and diagnostic testing systems for diseases such as TB have pivoted toward COVID-19, giving SSA an advantage over countries without existing infectious contagious diseases. Countries such as the United States should study how several low-income countries with fewer resources are able to prevent and control the spread of COVID-19 in efficient and cost-effective ways.

Limitations

Our data are available at the country level, meaning reported observed cases and rates of transmission are the country's average. Additional data collection and analysis are necessary to understand variations of COVID-19 transmission within countries. In addition, due to limitations in some country's health care infrastructure, several days of new COVID-19 infections and deaths may have been reported on a single day. To that end, we reported average daily cases to control for batch reports.

Conclusion

Public health surveillance systems are necessary to guide our leaders during disease outbreaks. Unfortunately, traditional surveillance metrics have undercounts, errors, and other data contamination that this study overcomes by using dynamic panel data and generalized method of the moments. Our methods control for data limitations and incomplete case ascertainment that exist in traditional surveillance systems.

Standard surveillance system metrics such as the prevalence and cumulative numbers of COVID-19–related morbidity and mortality are helpful; however, public health surveillance systems are limited in that they pick up the most severe cases. Incomplete case ascertainment is likely more pronounced during the COVID-19 pandemic where significant portions of the world's population infected with the novel coronavirus remain asymptomatic or presymptomatic. Even in a perfect world where surveillance metrics can enumerate all novel cases and deaths, important information is needed to guide mitigation and prevention policies. This study addresses that dearth by applying novel surveillance indicators including speed, acceleration or deceleration, jerk, and persistence of cases at a 7-day persistence time that measures the number of current new cases as a function of the number of new cases 7 days ago. Knowing the number of new cases or the total number of COVID-19 cases can be misleading. It is possible to have fewer total cases and be at risk for exponential growth and greater population health impact if the speed of the COVID-19 transmission is positive and if the speed of the COVID-19 transmission is accelerating, and moreover, if the speed is accelerating and that rate has positive jerks after subtracting out the acceleration rate from the prior week. It is critical that we understand how COVID-19 is spreading, and our study provides useful metrics to accomplish better surveillance in SSA.

Conflicts of Interest

None declared.

XSL•FO

RenderX

Multimedia Appendix 1
Sub-Saharan Africa COVID-19 trends.
[DOCX File , 2362 KB - jmir_v22i11e24248_app1.docx]

References

1. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol 2020 Jun;92(6):568-576 [FREE Full text] [doi: 10.1002/jmv.25748] [Medline: 32134116]
2. Yuen K, Ye Z, Fung S, Chan C, Jin D. SARS-CoV-2 and COVID-19: the most important research questions. Cell Biosci 2020;10:40. [doi: 10.1186/s13578-020-00404-4] [Medline: 32190290]
3. Kumari T, Shukla V. COVID-19: towards confronting an unprecedented pandemic. Int J Biol Innovations 2020;02(01):01-10. [doi: 10.46505/ijbi.2020.2101]
4. 7.2 human geography of Subsaharan Africa. University of Minnesota Libraries. 2012. URL: https://open.lib.umn.edu/worldgeography/chapter/7-2-human-geography-of-subsaharan-africa/
5. Diop BZ, Ngom M, Pougué Biyong C, Pougué Biyong JN. The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: a modelling study. BMJ Glob Health 2020 May;5(5):e002699. [doi: 10.1136/bmjgh-2020-002699] [Medline: 32451367]
6. Ehanire O. First case of corona virus confirmed in Nigeria. Nigeria Centre for Disease Control. 2020 Feb 28. URL: https://ncdc.gov.ng/news/227/first%2010case%2010of%2010corona%2010virus%2010disease%2010confirmed%2010in%2010nigeria) [accessed 2020-08-09]
7. Talisuna AO, Okiro EA, Yahaya AA, Stephen M, Bonkoungou B, Musa EO, et al. Spatial and temporal distribution of infectious disease epidemics, disasters and other potential public health emergencies in the World Health Organisation Africa region, 2016-2018. Global Health 2020 Jan 15;16(1):9 [FREE Full text] [doi: 10.1186/s12992-019-0540-4] [Medline: 31941554]
8. Ayebare R, Waitt P, Okello S, Kayiira M, Atim Ajok M, Nakatudde I, et al. Leveraging investments in Ebola preparedness for COVID-19 in Sub-Saharan Africa. AAS Open Res 2020;3:3. [doi: 10.12688/aasopenres.13052.1] [Medline: 32500116]
9. Ondoa P, Kebede Y, Loembe MM, Bhiman JN, Tessema SK, Sow A, et al. COVID-19 testing in Africa: lessons learnt. Lancet Microbe 2020 Jul;1(3):e103-e104. [doi: 10.1016/s2666-5247(20)30068-9]
10. COVID-19 in Africa: protecting lives and economies. United Nations Economic Commission for Africa. 2020 Apr. URL: https://www.uneca.org/sites/default/files/PublicationFiles/eca_covid_report_en_24apr_web1.pdf
11. Ritchie H, Roser M. Age structure. Our World in Data. 2019. URL: https://ourworldindata.org/age-structure
12. Renzaho AMN. The need for the right socio-economic and cultural fit in the COVID-19 response in sub-Saharan Africa: examining demographic, economic political, health, and socio-cultural differentials in COVID-19 morbidity and mortality. Int J Environ Res Public Health 2020 May 15;17(10):3445 [FREE Full text] [doi: 10.3390/ijerph17103445] [Medline: 32429123]
13. Zimmermann P, Curtis N. Coronavirus infections in children including COVID-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children. Pediatr Infect Dis J 2020 May;39(5):355-368 [FREE Full text] [doi: 10.1097/INF.0000000000002660] [Medline: 32310621]
14. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation and Treatment of Coronavirus. Treasure Island, FL: StatPearls Publishing LLC; 2020.
15. Velavan TP, Meyer CG. The COVID‑19 epidemic. Trop Med Int Health 2020 Feb 16;25(3):278-280. [doi: 10.1111/tmi.13383]
16. He W, Goodkind D, Kowal P. An aging world: 2015. United States Census Bureau. 2016. URL: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf
17. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. J Travel Med 2020 May 18;27(3):taaa038 [FREE Full text] [doi: 10.1093/jtm/taaa038] [Medline: 32227186]
18. Oyelami LO, Saibu OM, Adekunle BS. Determinants of financial inclusion in sub-Sahara African countries. Covenant J Business Soc Sci 2017;8(2):104-116.
19. Sharma G, Volgman AS, Michos ED. Sex differences in mortality from COVID-19 pandemic: are men vulnerable and women protected? JACC Case Rep 2020 Jul 15;2(9):1407-1410 [FREE Full text] [doi: 10.1016/j.jaccas.2020.04.027] [Medline: 32373791]
20. Klein SL, Dhakal S, Ursin RL, Deshpande S, Sandberg K, Mauvais-Jarvis F. Biological sex impacts COVID-19 outcomes. PLoS Pathog 2020 Jun;16(6):e1008570 [FREE Full text] [doi: 10.1371/journal.ppat.1008570] [Medline: 32569293]
21. Jin J, Bai P, He W, Wu F, Liu X, Han D, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health 2020;8:152. [doi: 10.3389/fpubh.2020.00152] [Medline: 32411652]
22. Massinga Loembé M, Tshangela A, Salyer SJ, Varma JK, Ouma AEO, Nkengasong JN. COVID-19 in Africa: the spread and response. Nat Med 2020 Jul;26(7):999-1003. [doi: 10.1038/s41591-020-0961-x] [Medline: 32528154]

23. Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. World Health Organization. 2020 Mar 29. URL: Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations: scientific brief, 27 March 2020

24. Nakkazi E. Obstacles to COVID-19 control in east Africa. Lancet Infect Dis 2020 Jun;20(6):660. [doi: 10.1016/s1473-3099(20)30382-0]

25. Connors J, Levy J. COVID-19 and its implications for thrombosis and anticoagulation. Blood 2020 Jun 04;135(23):2033-2040 [FREE Full text] [doi: 10.1182/blood.2020006000] [Medline: 32339221]

26. Kapata N, Ihekweazu C, Ntoumi F, Raji T, Chanda-Kapata P, Mwaba P, et al. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic. Lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. Int J Infect Dis 2020 Apr;93:233-236 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.049] [Medline: 32119980]

27. Mukwege D, Cadière GB, Vandenberg O. COVID-19 response in sub-Saharan low-resource setting: healthcare soldiers need bullets. Am J Trop Med Hyg 2020 Aug;103(2):549-550 [FREE Full text] [doi: 10.4269/ajtmh.20-0543] [Medline: 32588809]

28. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]

29. Edwards N. Tanzania's mild response to COVID-19 and its implications for the 2020 elections. Atlantic Council. 2020 Mar 27. URL: https://www.atlanticcouncil.org/blogs/africasource/tanzanias-mild-response-to-covid-19-and-its-implications-for-the-2020-elections/ [accessed 2020-08-11]

30. Resnick D. Trust in science and in government plays a crucial role in COVID-19 response. International Food Policy Research Institute. 2020. URL: https://www.ifpri.org/blog/trust-science-and-government-plays-crucial-role-covid-19-response

31. Olewe D. Coronavirus in Africa: whipping, shooting and snooping. BBC News. 2020. URL: https://www.bbc.com/news/world-africa-52214740 [accessed 2020-08-11]

32. Pavlik M. A great and sudden change: the global political violence landscape before and after the COVID-19 pandemic. ACLED. 2020. URL: https://acleddata.com/2020/08/04/a-great-and-sudden-change-the-global-political-violence-landscape-before-and-after-the-covid-19-pandemic/

33. Kennedy A, Resnick D. From farm to table: how are governments keeping food systems functioning during COVID-19? Int Food Policy Res Institute 2020:1-4. [doi: 10.2499/p15738coll2.133974]

34. Omoleke SA, Mohammed I, Saidu Y. Ebola viral disease in West Africa: a threat to global health, economy and political stability. J Public Health Afr 2016 Aug 17;7(1):534. [doi: 10.4081/jphia.2016.534] [Medline: 28299152]

35. Taub A. A new Covid-19 crisis: domestic abuse rises worldwide. The New York Times. 2020. URL: https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html

36. Nuwagira E, Muzoora C. Is sub-Saharan Africa prepared for COVID-19? Trop Med Health 2020;48:18 [FREE Full text] [doi: 10.1186/s41182-020-00206-x] [Medline: 32288543]

37. Kamal R, Kurani N, McDermott D, Cox X. How prepared is the US to respond to COVID-19 relative to other countries? Peterson-KFF Health System Tracker. 2020. URL: https://www.healthsystemtracker.org/chart-collection/how-prepared-is-the-us-to-respond-to-covid-19-relative-to-other-countries/#item-start [accessed 2020-09-21]

38. Dzinamarira T, Dzobo M, Chitungo I. COVID-19: a perspective on Africa's capacity and response. J Med Virol 2020 Jun 11:1 [FREE Full text] [doi: 10.1002/jmv.26159] [Medline: 32525568]

39. Rutter P, Hinman A, Hegg L, King D, Sosler S, Swezy V, et al. Transition planning for after polio eradication. J Infect Dis 2017 Jul 01;216(suppl_1):S287-S292 [FREE Full text] [doi: 10.1093/infdis/jix026] [Medline: 28838183]

40. Namubiru L, Wepukhulu KS. African governments warned of lawsuits for maternal deaths under COVID-19 lockdowns. Open Democracy. 2020. URL: https://www.opendemocracy.net/en/5050/african-governments-warned-lawsuits-maternal-deaths-covid-lockdowns/

41. Yao H, Chen J, Xu Y. Patients with mental health disorders in the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):e21. [doi: 10.1016/s2215-0366(20)30090-0]

42. Zeufack AG, Calderon C, Kambou G, Djiofack CZ, Megumi K, Vijdan K. Africa's Pulse, No. 21, Spring 2020: An Analysis of Issues Shaping Africa's Economic Future. Washington, DC: World Bank; 2020.

43. Thurlow J. 4. COVID-19 lockdowns have imposed substantial economic costs on countries in Africa. Int Food Policy Res Institute 2020:23-25. [doi: 10.2499/p15738coll2.133762_04]

44. Mental health and psychosocial considerations during the COVID-19 outbreak. World Health Organization. 2020 Mar 18. URL: https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf

45. Bouët A, Laborde D. 15. COVID-19 border policies create problems for African trade and economic pain for communities. Int Food Policy Res Institute 2020:69-72. [doi: 10.2499/p15738coll2.133762_15]

46. Hale T, Petherick A, Kira B. Coronavirus government response tracker. Blavatnik School of Government. 2020. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker [accessed 2020-08-30]

47. Amewu S, Asante S, Pauw K, Thurlow J. The economic costs of COVID-19 in sub-Saharan Africa: insights from a simulation exercise for Ghana. Int Food Policy Res Institute 2020:1-24. [doi: 10.2499/p15738coll2.133760]

48. Gostin LO, Hodge JG, Wiley LF. Presidential powers and response to COVID-19. JAMA 2020 Apr 28;323(16):1547-1548. [doi: 10.1001/jama.2020.4335] [Medline: 32186661]

49. Swinnen J. 3. COVID-19 is exacerbating inequalities in food security. Int Food Policy Res Institute 2020:20-22. [doi: 10.2499/p15738coll2.133762_03]

50. Hirvonen K, Abate GT, de Brauw A. 10. Survey suggests rising risk of food and nutrition insecurity in Addis Ababa, Ethiopia, as COVID-19 restrictions continue. Int Food Policy Res Institute 2020:46-49. [doi: 10.2499/p15738coll2.133762_10]

51. Reardon T, Bellemare MF, Zilberman D. 17. How COVID-19 may disrupt food supply chains in developing countries. Int Food Policy Res Institute 2020:78-80. [doi: 10.2499/p15738coll2.133762_17]

52. Measuring digital development: facts and figures 2019. ITU. 2020. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx#:~:text=ITU's%20Measuring%20digital%20development%3A%20Facts,gender%20gap%20is%20also%20growing

53. Garnier-Crussard A, Forestier E, Gilbert T, Krolak-Salmon P. Novel coronavirus (COVID-19) epidemic: what are the risks for older patients? J Am Geriatr Soc 2020 May;68(5):939-940 [FREE Full text] [doi: 10.1111/jgs.16407] [Medline: 32162679]

54. Peterman A, Potts A, O'Donnell M, Thompson K, Shah N, Oertelt-Prigione S, et al. Pandemics and violence against women and children. Center Global Dev 2020:1-43.

55. Afolabi MO, Folayan MO, Munung NS, Yakubu A, Ndow G, Jegede A, et al. Lessons from the Ebola epidemics and their applications for COVID-19 pandemic response in sub-Saharan Africa. Dev World Bioeth 2020 Jul 12:10.1111/dewb.12275 [FREE Full text] [doi: 10.1111/dewb.12275] [Medline: 32654261]

56. Sumner A, Hoy C, Ortiz-Juarez E. Estimates of the impact of COVID-19 on global poverty. UNU-WIDER 2020 Apr:800-809. [doi: 10.35188/unu-wider/2020/800-9]

57. Laurencin CT, McClinton A. The COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. J Racial Ethn Health Disparities 2020 Jun;7(3):398-402 [FREE Full text] [doi: 10.1007/s40615-020-00756-0] [Medline: 32306369]

58. Ozili PK. COVID-19 in Africa: socioeconomic impact, policy response and opportunities. Int J Sociol Soc Policy 2020:1-33. [doi: 10.2139/ssrn.3574767]

59. Wojcicki JM. The precious commodity of time and sub-Saharan Africa's success in keeping COVID-19 at bay. J Epidemiol Glob Health 2020 Sep;10(3):194-197. [doi: 10.2991/jegh.k.200727.001] [Medline: 32954707]

60. Kitara D, Ikoona E. Proposed strategies for easing COVID-19 lockdown measures in Africa. Pan Afr Med J 2020;36:179 [FREE Full text] [doi: 10.11604/pamj.2020.36.179.24194] [Medline: 32952823]

61. Mutsaka F. Progress against virus brings complacency in parts of Africa. Associated Press News. 2020. URL: https://apnews.com/article/virus-outbreak-harare-international-news-africa-zimbabwe-6c0d469d7acc0d8614921d8d9bf65318

62. Harris JE. Reopening under COVID-19: what to watch for. Natl Bureau Econ Res 2020:0898-2937. [doi: 10.3386/w27166]

63. Head T. Move to Level 2 confirmed: here are all the biggest lockdown changes. The South African. 2020. URL: https://www.thesouthafrican.com/news/what-are-new-level-2-regulations-lockdown-changes-south-africa-ramaphosa/

64. Hodge JG. Federal vs. state powers in rush to reopen amid the coronavirus pandemic. SSRN J 2020:1-4. [doi: 10.2139/ssrn.3587011]

65. Adalja AA, Toner E, Inglesby TV. Priorities for the US health community responding to COVID-19. JAMA 2020 Apr 14;323(14):1343-1344. [doi: 10.1001/jama.2020.3413] [Medline: 32125355]

66. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA 2020 Jun 23;323(24):2466-2467. [doi: 10.1001/jama.2020.8598] [Medline: 32391864]

67. Khunti K, Singh A, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? BMJ 2020 Apr 20;369:m1548. [doi: 10.1136/bmj.m1548] [Medline: 32312785]

68. Chowkwanyun M, Reed AL. Racial health disparities and COVID-19 — caution and context. N Engl J Med 2020 Jul 16;383(3):201-203. [doi: 10.1056/nejmp2012910]

69. Elflein J. Number of coronavirus (COVID-19) cases worldwide as of August 28, 2020, by country. Statista. 2020. URL: https://www.statista.com/statistics/1043366/novel-coronavirus-2019ncov-cases-worldwide-by-country/ [accessed 2020-08-28]

70. SARS-CoV-2 test tracker. FIND. 2020. URL: https://www.finddx.org/covid-19/test-tracker/ [accessed 2020-09-09]

71. FINDCov19TrackerData. GitHub. 2020. URL: https://github.com/dsbbfinddx/FINDCov19TrackerData

72. Sigmund M, Ferstl R. Panel vector autoregression in R with the package panelvar. Q Rev Economics Finance 2019 Jan:1. [doi: 10.1016/j.qref.2019.01.001]

73. Croissant Y, Millo G, Tappe K, Toomet O, Kleiber C, Zeileis A, et al. Linear models for panel data. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/web/packages/plm/plm.pdf [accessed 2020-08-31]

74. Oehmke J, Oehmke T, Singh L, Post L. Dynamic panel estimate-based health surveillance of SARS-CoV-2 infection rates to inform public health policy: model development and validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]

75. Oehmke J, Moss C, Singh L, Oehmke T, Post L. Dynamic panel surveillance of COVID-19 transmission in the United States to inform health policy: observational statistical study. J Med Internet Res 2020 Oct 05;22(10):e21955 [FREE Full text] [doi: 10.2196/21955] [Medline: 32924962]

XSL•FO

RenderX

Abbreviations

Africa CDC: Africa Centres for Disease Control and Prevention
SSA: sub-Saharan Africa
TB: tuberculosis
WHO: World Health Organization

XSL·FO
RenderX

<u>Original Paper</u>

Use of Facebook by Academic Medical Centers in Taiwan During the COVID-19 Pandemic: Observational Study

Wei-Min Chu[1,2,3,4,5], MD, PhD; Gow-Jen Shieh[6], PhD; Shi-Liang Wu[6], MS; Wayne Huey-Herng Sheu[4,6,7,8,9], MD, PhD

[1]Department of Family Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[2]Department of Occupational Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[3]Institute of Health Policy and Management, National Taiwan University, Taipei, Taiwan

[4]School of Medicine, National Yang-Ming University, Taipei, Taiwan

[5]School of Medicine, Chung Shan Medical University, Taichung, Taiwan

[6]Department of Top Hospital Administration, Taichung Veterans General Hospital, Taichung, Taiwan

[7]Division of Endocrinology & Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[8]School of Medicine, National Defense Medical Center, Taichung, Taiwan

[9]Institute of Medical Technology, College of Life Science, National Chung-Hsing University, Taichung, Taiwan

Corresponding Author:
Wayne Huey-Herng Sheu, MD, PhD
Department of Top Hospital Administration
Taichung Veterans General Hospital
1650 Taiwan Boulevard Section 4
Taichung, 40705
Taiwan
Phone: 886 23592525 ext 2005
Email: whhsheu@vghtc.gov.tw

Abstract

Background: The battle against COVID-19 remains ongoing, and social media has played an important role during the crisis for both communication and health promotion, particularly for health care organizations. Taiwan's success during the COVID-19 outbreak is well known and the use of social media is one of the key contributing factors to that success.

Objective: This nationwide observational study in Taiwan aimed to explore the use of Facebook by academic medical centers during the COVID-19 pandemic.

Methods: We conducted a nationwide observational study of all Facebook fan page posts culled from the official accounts of all medical centers in Taiwan from December 2019 to April 2020. All Facebook posts were categorized into either COVID-19–related posts or non–COVID-19–related posts. COVID-19–related posts were split into 4 categories: policy of Taiwan's Center for Disease Control (TCDC), gratitude notes, news and regulations from hospitals, and education. Data from each post was also recorded as follows: date of post, headline, number of "likes," number of messages left, number of shares, video or non-video post, and date of search.

Results: The Facebook fan pages of 13 academic medical centers, with a total of 1816 posts, were analyzed. From January 2020, the percentage of COVID-19 posts increased rapidly, from 21% (January 2020) to 56.3% (April 2020). The trends of cumulative COVID-19 posts and reported confirmed cases were significantly related (Pearson correlation coefficient=0.93, *P*<.001). Pages from private hospitals had more COVID-19 posts (362 versus 289), as well as more video posts (72 posts, 19.9% versus 36 posts, 12.5%, *P*=.011), when compared to public hospitals. However, Facebook pages from public hospitals had significantly more "likes," comments, and shares per post (314, 5, 14, respectively, *P*<.001). Additionally, medical centers from different regions displayed different strategies for using video posts on Facebook.

Conclusions: Social media has been a useful tool for communication during the COVID-19 pandemic. This nationwide observational study has helped demonstrate the value of Facebook for academic medical centers in Taiwan, along with its engagement efficacy. We believe that the experience of Taiwan and the knowledge it can share will be helpful to health care organizations worldwide during our global battle against COVID-19.

Introduction

Background

COVID-19 originated from Wuhan, China, and has become a worldwide pandemic with more than 4 million cases and more than 200,000 deaths globally as of May 2020 [1-3]. Following Wuhan's lockdown [4], most countries with severe outbreaks took multiple measures to contain the virus and stop it from spreading. Steps that were taken included limiting both international and domestic flights, enhancing border controls, and emphasizing the importance of wearing masks and comprehensive hand washing. However, without a much-needed vaccine, the pandemic remains ongoing.

Many scientists have compared COVID-19 with the severe acute respiratory syndrome (SARS) outbreak in 2003 [5]. However, the world we now live in is so extremely different from the world we remember during the 2003 SARS epidemic, as social media now plays such a big role in our lives, more so than any other time in history. Social media has been a valuable tool for more than a decade for both health care organizations and health care professionals, and enables them to connect with people at risk, increase the health literacy of the general population, and improve health outcomes [6-8].

Prior Work

Thus, as the COVID-19 outbreak unfolded, the characteristics and capabilities of social media have become more evident. Social media directs people to trusted sources, counteracts misinformation, enables connectivity and psychological first aid, advances remote learning, and even accelerates research [9]. Moreover, social media helps to publicly reveal any early warning signals once an outbreak starts [10]. Additionally, the data coming from internet research surrounding social media helps medical professionals and scientists predict outbreaks [11].

Among all available social media, Facebook plays a special role. Reportedly, Facebook has more than 180 million users in the United States and Canada, with that number still increasing [12]. Facebook has not only been used as a communication tool for health care issues, but also as a tool for recruiting participants into medical research studies [13]. During the previous Zika outbreak, Facebook proved that it can be a cost-effective platform for engaging more people to help ensure better disease prevention [14].

Role of Taiwan Against COVID-19

Taiwan is one of the many countries with a large number of Facebook users [15]. Taiwan's experience in the fight against COVID-19 is unique, as its low number of cases and fatalities are a result of proactive measures, new technology, and big data analytics [16-18]. Taiwan's Center for Disease Control (TCDC) used social media widely, mainly the LINE app and Facebook,

to maintain surveillance and communicate risk to the public [19].

Research Gap

Although a previous study showed that social media can be an effective tool for engagement [20], it is not yet clear what type of posts grab more public attention. Similarly, what post timing is the most appropriate to have more engagement such as likes, comments, and shares? In addition, previous studies have shown that video posts had a significant association with favorable engagement rates [12], but how were video posts used on social media by health authorities during a pandemic such as COVID-19 and what were the effects? There is still a research gap when it comes to social media, information dissemination, and disease outbreaks.

There has been no study that has focused on how health care organizations have used social media to communicate with the public and increase health literacy. In Taiwan, medical centers are academic medical facilities providing the highest level of medical care as accredited by the Joint Commission of Taiwan (JCT), under the supervision of the Ministry of Health and Welfare of Taiwan (MOHW). Different from the role of TCDC and the MOHW, medical centers not only educate people for health promotion, but also treat people with illnesses directly. During the COVID-19 pandemic, medical centers in Taiwan have the most responsibility to test and receive patients with COVID-19. Such a "dual role" made medical centers uniquely have bidirectional communication with the public during the pandemic [21].

From previous studies, we understood that the use of Facebook is helpful for raising public awareness and sharing information rapidly and has "listening" characteristics during public health emergencies, such as the Zika outbreak [22]. In addition, a study focused on Facebook use during an Ebola outbreak suggested that information behavior and audience engagement was topic-dependent and regular health promotion efforts should continue taking place during a health crisis [20]. However, we still want to know whether this kind of observation persists in a different context. Does regular health promotion continue taking place during health emergencies? Analyzing the use of Facebook during the COVID-19 pandemic gave us the opportunity to understand the role of social media and the response of health care organizations more clearly.

Goal of This Study

We believe that social media has played an important role during the worldwide COVID-19 crisis. However, no study has focused on how health care organizations have used social media to communicate with the public and increase health literacy. In a previous study, medical centers in Taiwan ran their social media better than regional hospitals and local community hospitals; thus, we conducted this study as a nationwide observational

study involving the use of Facebook data from academic medical centers in Taiwan during the COVID-19 outbreak period.

Methods

Data Sources

We conducted a nationwide observational study of all Facebook fan page posts from official accounts linked to all medical centers in Taiwan during the period spanning December 2019 to April 2020. For comparison, we also surveyed all Facebook posts by the MOHW and TCDC.

Study Group Identification and Data Extraction

We first checked the official names of the medical centers in Taiwan on the JCT website. At the end of 2019, there were 19 medical centers in Taiwan. All Facebook fan page posts by the medical centers, MOHW, and TCDC were created by a branding team from Taichung Veterans General Hospital (TCVGH). The hospital team was established in 2016 and is led by the superintendent of the hospital. The authors searched all Facebook fan pages of the medical centers during the study period by using the names of the hospitals or their abbreviations. If the medical center did not have a Facebook fan page account, it was excluded from the study. We also recorded whether a medical center was a public or private hospital, as well as their location in Taiwan.

Research Variables

All Facebook posts were categorized into either COVID-19– or non–COVID-19–related posts. A modified focus group reviewed all posts and divided them according to their content characteristics into 14 unique themes according to a previously published study protocol [23]. We then classified the 14 themes into 4 groups for analysis: policy of TCDC, gratitude notes, news/regulations of hospitals, and education. The classification was decided by a modified focus group. A senior leader of our team was appointed as "gatekeeper" of our classification and she would review all posts and their categories to make sure that no classification bias happened. Data regarding each post was also recorded as follows: date of post, headline, number of "likes," number of messages left, number of shares, video or no-video post, and date of search. Additionally, the number of confirmed cases of COVID-19 was recorded after being taken from public data made available by TCDC. From a previous study, Facebook "Likes" can be used as an indicator of hospital quality [24].

We also categorized all medical centers into different regions and public or private sector. Public hospitals and private hospitals have very different resources in Taiwan. The public sector receives relatively fewer resources from the government despite bearing a heavier burden than the private sector. Personnel and purchasing systems in public institutions are also less flexible, making them less competitive than private hospitals in providing medical services. Thus, better use of social media to promote the hospital itself became more important for public hospitals. That is why we separated all medical centers into public or private classifications. As for the regions, health care in Taiwan is managed centrally by the Bureau of National Health Insurance (BNHI). The divisions of the BNHI are according to region (eg, Northern, Central, Southern, and Eastern) because medical resources are different from one region to another due to geographical differences.

Statistical Analysis

For the purpose of descriptive analysis, we used the chi-square test to compare the categorized variables. The Kolmogorov-Smirnov test was used to test normality. If the data was nonparametric, the Mann-Whitney U test and Kruskal-Wallis test were used to compare the continuous and categorized variables, respectively. Post hoc analysis with the Dunn post hoc test was performed if there was statistical significance. Pearson correlation coefficients were used to analyze trends of the cumulative COVID-19 posts and cumulative confirmed cases. We used the chi-square test to compare the percentage of video posts among all the medical centers, while Bonferroni correction post hoc tests were performed if there was statistical significance. Statistical significance was set at $P<.05$. All data were analyzed using SPSS (Version 22.0, IBM).

Results

All Facebook posts from 13 academic medical centers during the period between December 1, 2019, and April 30, 2020, were analyzed. Table 1 shows the scale of the surveyed medical centers during the research period and their Facebook fan page status. In mid-2020, there were 19 medical centers in Taiwan, and among them, 6 were public hospitals, while the others were private. Of them, only 4 medical centers had not established a Facebook fan page. There were 2 other medical centers whose Facebook fan pages had not been renewed for years. Thus, we analyzed a total of 13 medical centers and their Facebook fan page posts during the COVID-19 pandemic. Chung Shan Medical University Hospital had the most COVID-19 posts (95 posts, 14.6%) among all medical centers.

Table 1. Nationwide medical center information and Facebook fan pages in Taiwan.

Region, ownership, and name of hospital	Hospital scale[a]		Creation date of Facebook fan page	Number of Facebook fan page "Likes"[b]	Number of COVID-19 Facebook posts, n (%)[c]
	Number of beds	Number of employees			
Northern, Public					
National Taiwan University Hospital	2289	6700	July 8, 2010	10,107	25 (3.8)
Tri-Service General Hospital	1793	3000	June 30, 2013	13,702	65 (10.0)
Taipei Veterans General Hospital	2803	6655	February 18, 2014	3098	4 (0.6)
Northern, Private					
Chang Gung Memorial Hospital, Linkou	3700	20,000	—[d]	—	0 (0.0)
Cathay General Hospital	816	3400	—[d]	—	0 (0.0)
MacKay Memorial Hospital	938	3600	August 4, 2016	3228	5 (0.8)
Shin Kong Wu Ho-Su Memorial Hospital	833	2300	March 10, 2020	212	57 (8.8)
Taipei Municipal Wanfang Hospital	726	1696	January 12, 2018	1676	15 (2.3)
Far Eastern Memorial Hospital	1383	3369	February 10, 2012	2106	68 (10.4)
Central, Public					
Taichung Veterans General Hospital	1527	4000	January 30, 2012	33,343	66 (10.1)
Central, Private					
Chung Shan Medical University Hospital	1003	2300	March 30, 2017	4945	95 (14.6)
China Medical University Hospital[e]	1722	4000	April 3, 2012	389	0 (0.0)
Changhua Christian Hospital	1244	5011	—[d]	—	0 (0.0)
Southern, Public					
National Cheng Kung University Hospital	1189	5035	July 9, 2015	11,891	63 (9.7)
Kaohsiung Veterans General Hospital	1482	3561	September 10, 2013	17,778	66 (10.1)
Southern, Private					
Chi Mei Medical Center	1278	3664	—[d]	—	0 (0.0)
Kaohsiung Chang Gung Memorial Hospital[f]	2691	6200	February 16, 2011	7398	0 (0.0)
Kaohsiung Medical University Chung-Ho Memorial Hospital	1720	4078	November 25, 2019	1548	51 (7.8)
Eastern, Private					
Hualien Tzu Chi Hospital	947	1987	August 20, 2015	8577	71 (10.9)

[a]The data was obtained from the official website of the hospital (access date: April 30, 2020).

[b]The data was obtained from the Facebook fan page of the hospital (access date: April 30, 2020).

[c]The collection period of COVID-19 Facebook posts was from December 31, 2019, to April 30, 2020. Percentages were calculated by taking the number of COVID-19 posts by an individual medical center (n) and dividing it by the total number (N) of COVID-19 posts of all national medical centers (multiplied by 100).

[d]Not available (the official Facebook fan pages were not found).

[e]Date of the most recent Facebook fan page post: November 2, 2012.

[f]Date of the most recent Facebook fan page post: November 15, 2013.

Figure 1 shows the cumulative COVID-19 posts per week from all academic medical centers, TCDC, and the MOWH, along with the cumulative confirmed cases of COVID-19 per week in Taiwan. There was a relationship between cumulative COVID-19 posts per week from medical centers and confirmed cases per week (Pearson correlation coefficient=0.56, P=.01). We also found a similar trend regarding posts from the MOHW (Pearson correlation coefficient=0.58, P=.01). However, the relationship between COVID-19 posts by TCDC and confirmed cases was relatively low (Pearson correlation coefficient=0.03,

XSL•FO
RenderX

P=.89). For most of the study period, medical centers had the most COVID-19–related posts per week; however, posts from TCDC took the lead just before the small outbreak that occurred around March 18, 2020. Multimedia Appendix 1 reveals the cumulative COVID-19 posts from academic medical centers, with the cumulative confirmed cases of COVID-19.

Figure 1. COVID-19 confirmed cases per week in Taiwan and Facebook posts per week by the MOHW, TCDC, and nationwide medical centers from December 2019 to April 2020. CECC: Central Epidemic Command Center; MOHW: Ministry of Health and Welfare; TCDC: Taiwan's Center for Disease Control.
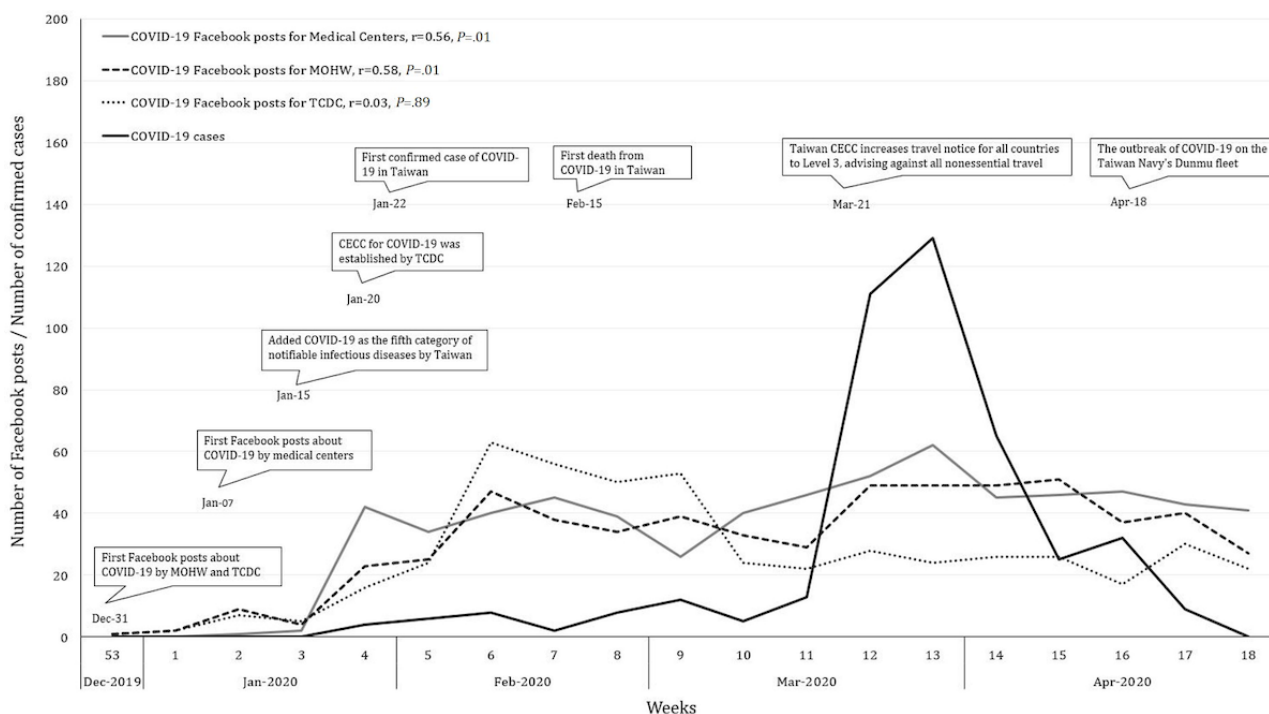
Table 2 shows the trends and distributions of COVID-19–related posts and their categories, along with non–COVID-19 posts. A total of 1816 posts were analyzed during the study period. There were no COVID-19 posts in December 2019 before the outbreak in Wuhan, China. From January 2020, COVID-19 posts increased rapidly, from 21% (January 2020) to 56.3% (April 2020) of total posts. The number of non–COVID-19 posts decreased significantly due to the crowding out effect. Within the 4 categories of COVID-19 posts, at the beginning of the pandemic, most were about National Health Command Center news and regulations (40%); however, as the outbreak continued in February and March, hospital-released news and policy announcements had become the main focus of posts (52% and 36%, respectively). In April, most COVID-19 posts expressed gratitude (42%). After logistic regression surrounding the trend of COVID-19 posts by month, the P value for a trend was <.001 (Multimedia Appendix 2). Compared with January, the odds ratio (OR) of COVID-19–related posts of nationwide medical centers was significantly increased in February, March, and April (Multimedia Appendix 3).

Table 2. COVID-19 and non–COVID-19 Facebook posts of nationwide medical centers from December 2019 to April 2020.

| Type of post	2019	2020				All	P value
	December, n (%)	January, n (%)	February, n (%)	March, n (%)	April, n (%)	Total, n (%)	
Total	407 (100.0)	357 (100.0)	308 (100.0)	373 (100.0)	371 (100.0)	1816 (100.0)	
Disease category							
Non–COVID-19 posts	407 (100.0)	282 (79.0)	158 (51.3)	156 (41.8)	162 (43.7)	1165 (64.2)	<.001
COVID-19 posts	0 (0.0)	75 (21.0)	150 (48.7)	217 (58.2)	209 (56.3)	651 (35.8)	
Topic category							
Policy of Taiwan's Center for Disease Control	0 (0.0)	30 (40.0)	27 (18.0)	44 (20.3)	23 (11.0)	124 (19.0)	<.001
Gratitude notes	0 (0.0)	0 (0.0)	5 (3.3)	56 (25.8)	88 (42.1)	149 (22.9)	
News and regulations from hospitals	0 (0.0)	29 (38.7)	78 (52.0)	80 (36.9)	65 (31.1)	252 (38.7)	
Education	0 (0.0)	16 (21.3)	40 (26.7)	37 (17.1)	33 (15.8)	126 (19.4)	

XSL•FO
RenderX

Multimedia Appendix 4 shows the engagement of all COVID-19 posts, including "likes" per post, shares per post, and comments per post. Private hospitals had more COVID-19 posts and more video posts (72 posts, 19.9%) during the research period, but public hospitals had significantly more "likes" (314), comments (5), and shares (14) per post. Regarding different regions, the northern regional medical centers had the most COVID-19 posts (239), while the central regional medical centers had the most "likes" (197) and comments (4) per post. The southern regional hospitals had the most shares per post (14). For video posts regarding COVID-19, the central public medical center had significantly more video posts compared to medical centers of every other region and type.

Figure 2 shows the trend of "likes" per COVID-19 post for both the different regions and private/public sectors. The central public medical centers had the most "likes" per post, while the northern private hospitals had the fewest "likes" per posts. Multimedia Appendix 5 outlines different post contents and their number of fan likes, comments, and shares, as well as the percentage of video posts published by the central government health agency and 13 nationwide medical centers before and after the COVID-19 outbreak on March 18, 2020. The number of likes and comments increased significantly after the outbreak for the MOHW, TCDC, and medical centers. The number of shares increased significantly after the outbreak only for TCDC and medical centers. We observed that TCDC used more video posts after the outbreak (*P*<.001). Medical centers had more education-related COVID-19 posts compared with TCDC and the MOHW, and those educational posts received more likes and comments after the outbreak. We also analyzed all 4 categories of COVID-19–related posts from each medical center and their effectiveness at achieving engagement (Multimedia Appendix 6).

Figure 2. The number of "likes" on COVID-19 Facebook posts by nationwide medical centers in Taiwan, by region and ownership.
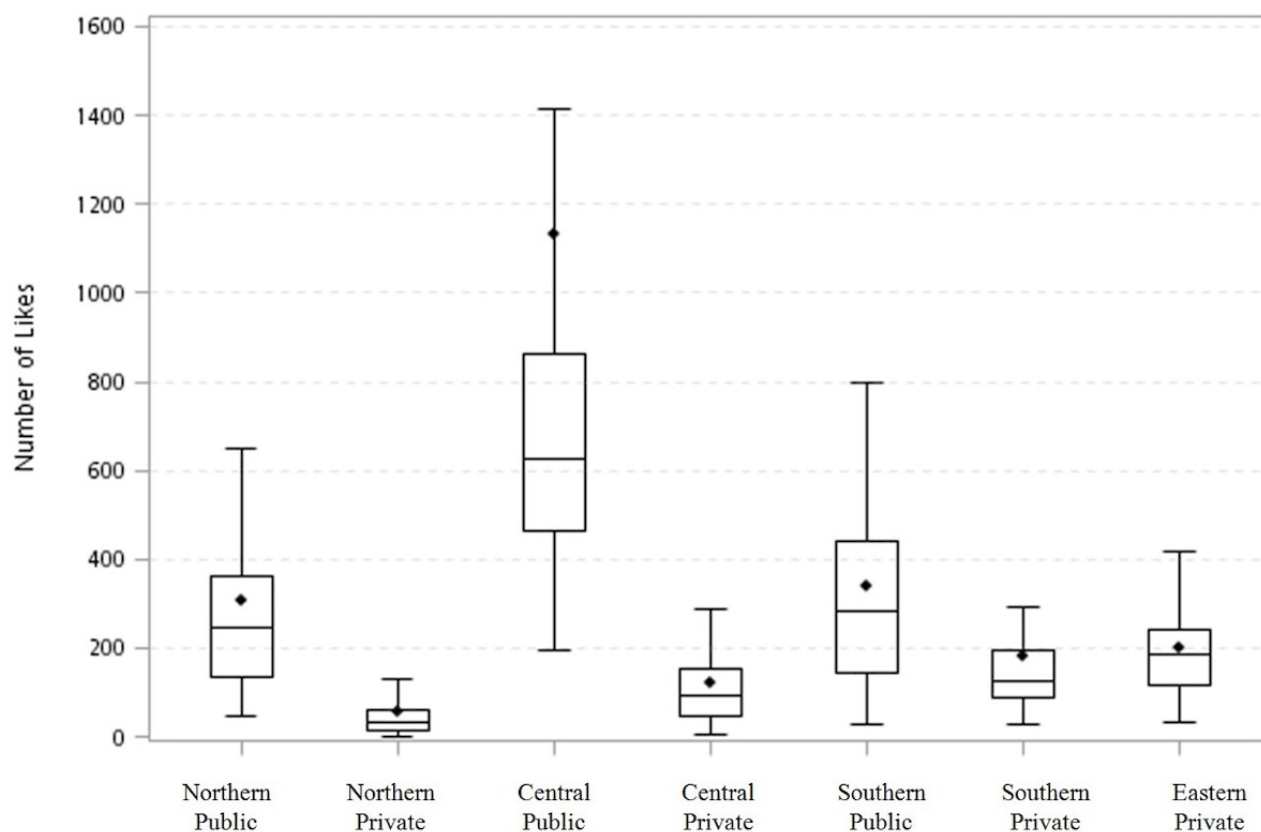
Discussion

Principal Results

To the best of our knowledge, this is the first nationwide study that has explored how academic medical centers in Taiwan used Facebook as a tool for the communication of risk, as well as to increase public health literacy after the outbreak of COVID-19. The main results are as follows:

1. COVID-19 Facebook posts per week increased, along with the scale of public awareness and cumulative cases per week of COVID-19.

2. Non–COVID-19 posts statistically decreased during the COVID-19 pandemic.
3. Public medical centers provided more engagement with the public (through "likes," comments, and shares), while private medical centers created video posts more often, particularly for the purpose of expressing gratitude.
4. Medical centers in different regions displayed different strategies when using video posts on Facebook.

Although a previous study demonstrated a relationship between social media and infectious disease outbreaks, there was no further analysis of post contents and outcomes. In our study, COVID-19–related posts per week were related to the number of confirmed cases per week, suggesting that medical centers

emphasized the disease during the outbreak. However, COVID-19–related posts by TCDC rose before the outbreak of COVID-19, suggesting that TCDC's response was fast and could be regarded as a "warning" during health crises. Furthermore, gratitude–related and government policy–related COVID-19 posts received many likes and comments, suggesting that these two kinds of posts got more attention. Our findings could be a guide for all medical facilities to have better engagement with the public when public health crises occur.

Comparison With Prior Work

Social media has been largely used by health care organizations to communicate with the public and improve public awareness surrounding health issues [25], particularly during disease outbreaks such as COVID-19 [26]. A previous study showed that Facebook has been widely used as a tool for both the promotion of health [27,28] and disease management [29,30]. In the United States, most hospitals use their Facebook account as a platform [31]. In addition, a 2017 study revealed that 51.1% (213/417) of hospitals in Taiwan had a Facebook fan page, with an average of 31 posts by the medical centers [24]. Our study offers similar results, with academic medical centers having a higher proportion (13/19, 68%) of Facebook use as a communication tool, and an average post count of 27.9 during the research period.

However, few studies have focused on Facebook use with regard to disease outbreak communication. One study in Malaysia explored Facebook use by health care authorities during the Zika outbreak, showing that health authorities posted updates most frequently within the first two weeks after the outbreak was declared. The findings from our study were different, as we found that the greatest number of COVID-19 posts was published in March 2020, more than one month after the first confirmed case in Taiwan; confirmed cases were increasing rapidly in March. This suggests that public awareness arises from a feeling of imminent threat, and not just from a public declaration made by health care authorities.

As the COVID-19 pandemic is now a worldwide health hazard, health care surge capacity remains an important issue that every country is concerned about [32]. In our study, altered social media content was also noticed, as the number of non–COVID-19 posts rapidly decreased from 100% in December 2019 to 43.7% in April 2020. This discovery implies that much content surrounding health promotion or medical education for chronic diseases such as hypertension or diabetes was not disseminated due to COVID-19 posts capturing the majority of public attention. The long-term effect of such an altered "social media capacity" should be further examined.

We found that medical centers from different regions and with different types of ownership implemented varying strategies when using video posts as a method of communication, with the percentage of video posts pertaining to COVID-19 ranging from 0% to 40.9%. Previous studies found that, when compared with other post types, video posts were the only type that had a significant association with favorable engagement rates, while posts linking shared video posts had a negative impact on engagement [7,33]. Another study revealed that both photos and videos were linked to an increase in the amount of time

from the most recent interaction activity [12]. Additionally, video post pairings involving text are easy to remember, and are more likely to capture the attention of the younger generation [34]. Health care organizations should use video posts more often to achieve better engagement and understanding.

Among all regions and types of ownership, one central and public hospital, Taichung Veterans General Hospital (TCVGH), experienced the most "likes" per post (more than 600 on average), and its percentage of video posts was the highest (40.9%).

TCVGH is one of the 19 medical centers in Taiwan and was established in 1982. It is among the medical facilities found in central Taiwan, and is the only government medical center providing medical services to the public. The branding team of VGHTC was first set up in 2016 and they started promoting the hospital branding in 2017; as of May 2020, TCVGH had the largest number of Facebook fans among all medical centers in Taiwan [23]. Its main mission was to establish branding strategies and integrate segmental branding messages. Under the leadership of the hospital's superintendent, 11 divisions were recruited to form this special task force. During the COVID-19 pandemic, top administrative officers held regular meetings every week and discussed social media strategies including videos, broadcasts, and touching stories to increase public engagement and health literacy. The following may be reasons for the effectiveness of TCVGH's approach: (1) Top-down leadership in forming the branding team, (2) a high number of Facebook fans who regularly visit the official Facebook page, and (3) various types of posts, such as those including video, to increase engagement. We believe that lessons can be learned from the above experiences.

Limitations

This study has several limitations. First, we only analyzed data originating from Facebook posts taken from all nationwide medical centers without analyzing any other social media platforms, such as Twitter, Instagram, or LINE. However, Facebook remains the major social media platform with which the Taiwanese people are most familiar. Second, this study was a cross-sectional study, and the content on social media changes every day, so the number of posts or engagement efficacy could have changed by the time the research team accessed the posts. Third, we only enrolled official Facebook fan pages in the analysis. However, we found that some departments within the medical centers had their own Facebook fan page, which usually focused on specific organ systems or diseases. Fourth, classification errors could have occurred because there were 4 types of post category, and it is possible that a given post contained more than one issue. However, classification errors were likely rare because each post categorization was decided by a modified focus group, while a senior division leader served as "gatekeeper" when reviewing the process. Fifth, some data was not made publicly available, such as post frequency, number of views per post, and the changing number of fans at each medical center. That data would improve our understanding of the true effectiveness of various posting strategies. Therefore, further research is warranted to better understand the relationship between dynamic posts and the health outcomes of the people

engaged. In addition, some medical centers have more than one official Facebook page; thus, further analysis of multiple Facebook pages within a single hospital may help determine the effectiveness of different Facebook pages set up by different divisions.

Conclusion

Social media remains an invaluable tool for communication during the COVID-19 pandemic. This nationwide observational study has demonstrated the use of Facebook by academic medical centers in Taiwan and its engagement efficacy. Taiwan continues to play a unique role in the battle against COVID-19, and we would therefore like to share our experience in the use of social media for communication and health promotion, as we believe our knowledge will be helpful to health care organizations worldwide.

Multimedia Appendix 1
COVID-19 Facebook posts of nationwide medical centers and COVID-19 confirmed cases from December 2019 to April 2020 in Taiwan.
[PNG File , 318 KB - jmir_v22i11e21501_app1.png]

Multimedia Appendix 2
The percentage of COVID-19–related Facebook posts by nationwide medical centers from December 2019 to April 2020 in Taiwan.
[PNG File , 195 KB - jmir_v22i11e21501_app2.png]

Multimedia Appendix 3
The odds ratio of COVID-19–related Facebook posts by nationwide medical centers from January to April 2020 in Taiwan.
[PNG File , 61 KB - jmir_v22i11e21501_app3.png]

Multimedia Appendix 4
The relationship of COVID-19 Facebook posts and ownership/region of nationwide medical centers in Taiwan from January to April 2020.
[DOCX File , 21 KB - jmir_v22i11e21501_app4.docx]

Multimedia Appendix 5
The number of fan likes, comments, and shares on Facebook fan pages, and the percentage of video posts by the central government health agency and 13 nationwide medical centers.
[DOCX File , 21 KB - jmir_v22i11e21501_app5.docx]

Multimedia Appendix 6
The number of fan likes, comments, and shares on Facebook fan pages, as well as the percentage of video posts published by the central government health agency and 13 nationwide medical centers before and after the COVID-19 outbreak on March 18, 2020.
[DOCX File , 62 KB - jmir_v22i11e21501_app6.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]

XSL•FO
RenderX

2. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report 115. 2020 May 14. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200514-covid-19-sitrep-115.pdf?sfvrsn=3fce8d3c_6 [accessed 2020-05-15]

3. Huang W, Teng L, Yeh T, Chen Y, Lo W, Wu M, et al. 2019 novel coronavirus disease (COVID-19) in Taiwan: Reports of two cases from Wuhan, China. J Microbiol Immunol Infect 2020 Jun;53(3):481-484 [FREE Full text] [doi: 10.1016/j.jmii.2020.02.009] [Medline: 32111449]

4. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):1 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]

5. Yang Y, Peng F, Wang R, Yang M, Guan K, Jiang T, et al. Corrigendum to "The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China" [J. Autoimmun. 109C (2020) 102434]. J Autoimmun 2020 Jul;111:102487 [FREE Full text] [doi: 10.1016/j.jaut.2020.102487] [Medline: 32423578]

6. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]

7. Lister C, Royne M, Payne HE, Cannon B, Hanson C, Barnes M. The Laugh Model: Reframing and Rebranding Public Health Through Social Media. Am J Public Health 2015 Nov;105(11):2245-2251. [doi: 10.2105/AJPH.2015.302669] [Medline: 26378824]

8. Novillo-Ortiz D, Hernández-Pérez T. Social media in public health: an analysis of national health authorities and leading causes of death in Spanish-speaking Latin American and Caribbean countries. BMC Med Inform Decis Mak 2017 Feb 03;17(1):16 [FREE Full text] [doi: 10.1186/s12911-017-0411-y] [Medline: 28158986]

9. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]

10. Zhu Y, Fu K, Grépin KA, Liang H, Fung IC. Limited Early Warnings and Public Attention to Coronavirus Disease 2019 in China, January-February, 2020: A Longitudinal Cohort of Randomly Sampled Weibo Users. Disaster Med Public Health Prep 2020 Apr 03:1-4 [FREE Full text] [doi: 10.1017/dmp.2020.68] [Medline: 32241328]

11. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]

12. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49 [FREE Full text] [doi: 10.1186/s12911-017-0447-z] [Medline: 28431582]

13. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

14. Lwin MO, Lu J, Sheldenkar A, Schulz PJ. Strategic Uses of Facebook in Zika Outbreak Communication: Implications for the Crisis and Emergency Risk Communication Model. Int J Environ Res Public Health 2018 Sep 10;15(9):1 [FREE Full text] [doi: 10.3390/ijerph15091974] [Medline: 30201929]

15. Internet World Stats. Internet Usage in Asia. URL: https://www.internetworldstats.com/stats3.htm [accessed 2020-11-12]

16. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]

17. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. JAMA Intern Med 2020 Sep 01;180(9):1156-1163 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]

18. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]

19. Cheng H, Li S, Yang C. Initial rapid and proactive response for the COVID-19 outbreak - Taiwan's experience. J Formos Med Assoc 2020 Apr;119(4):771-773 [FREE Full text] [doi: 10.1016/j.jfma.2020.03.007] [Medline: 32222336]

20. Strekalova YA. Health Risk Information Engagement and Amplification on Social Media. Health Educ Behav 2017 Apr;44(2):332-339. [doi: 10.1177/1090198116660310] [Medline: 27413028]

21. Kowitt SD, Schmidt AM, Hannan A, Goldstein AO. Awareness and trust of the FDA and CDC: Results from a national sample of US adults and adolescents. PLoS One 2017;12(5):e0177546 [FREE Full text] [doi: 10.1371/journal.pone.0177546] [Medline: 28520750]

22. Vijaykumar S, Meurzec RW, Jayasundar K, Pagliari C, Fernandopulle Y. What's buzzing on your feed? Health authorities' use of Facebook to combat Zika in Singapore. J Am Med Inform Assoc 2017 Nov 01;24(6):1155-1159. [doi: 10.1093/jamia/ocx028] [Medline: 28449047]

23. Shieh G, Wu S, Tsai C, Chang C, Chang T, Lui P, et al. A Strategic Imperative for Promoting Hospital Branding: Analysis of Outcome Indicators. Interact J Med Res 2020 Jan 22;9(1):e14546 [FREE Full text] [doi: 10.2196/14546] [Medline: 32012047]

24. Yang P, Lee W, Liu H, Shih M, Chen T, Chou L, et al. Use of Facebook by Hospitals in Taiwan: A Nationwide Survey. Int J Environ Res Public Health 2018 Jun 06;15(6):1 [FREE Full text] [doi: 10.3390/ijerph15061188] [Medline: 29882819]

25. Thaker SI, Nowacki AS, Mehta NB, Edwards AR. How U.S. hospitals use social media. Ann Intern Med 2011 May 17;154(10):707-708. [doi: 10.7326/0003-4819-154-10-201105170-00021] [Medline: 21576547]

26. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]

27. Kousoulis AA, Kympouropoulos SP, Pouli DK, Economopoulos KP, Vardavas CI. From the Classroom to Facebook: A Fresh Approach for Youth Tobacco Prevention. Am J Health Promot 2016 May;30(5):390-393. [doi: 10.1177/0890117116646345] [Medline: 27404648]

28. Cheung YTD, Chan CHH, Wang MP, Li HCW, Lam T. Online Social Support for the Prevention of Smoking Relapse: A Content Analysis of the WhatsApp and Facebook Social Groups. Telemed J E Health 2017 Jun;23(6):507-516. [doi: 10.1089/tmj.2016.0176] [Medline: 27911654]

29. Kim C, Kang BS, Choi HJ, Lee YJ, Kang GH, Choi WJ, et al. Nationwide online social networking for cardiovascular care in Korea using Facebook. J Am Med Inform Assoc 2014;21(1):17-22 [FREE Full text] [doi: 10.1136/amiajnl-2012-001465] [Medline: 23703828]

30. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]

31. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]

32. Verelst F, Kuylen E, Beutels P. Indications for healthcare surge capacity in European countries facing an exponential increase in coronavirus disease (COVID-19) cases, March 2020. Euro Surveill 2020 Apr;25(13):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000323] [Medline: 32265003]

33. A Rahim AI, Ibrahim MI, A Salim FN, Ariffin MAI. Health Information Engagement Factors in Malaysia: A Content Analysis of Facebook Use by the Ministry of Health in 2016 and 2017. Int J Environ Res Public Health 2019 Feb 18;16(4):1 [FREE Full text] [doi: 10.3390/ijerph16040591] [Medline: 30781691]

34. Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020 Mar 25;368:m1160. [doi: 10.1136/bmj.m1160] [Medline: 32213480]

Abbreviations

SARS: severe acute respiratory syndrome
TCDC: Taiwan's Center for Disease Control
JCT: Joint Commission of Taiwan
MOHW: Ministry of Health and Welfare
TCVGH: Taichung Veterans General Hospital

Original Paper

Implementation of a Self-Triage Web Application for Suspected COVID-19 and Its Impact on Emergency Call Centers: Observational Study

Simon Galmiche[1]; Eve Rahbe[1]; Arnaud Fontanet[1,2], MD, DrPH; Aurélien Dinh[3], MD; François Bénézit[4], MD; François-Xavier Lescure[5,6], MD, PhD; Fabrice Denis[7], MD, PhD

[1]Emerging Diseases Epidemiology Unit, Institut Pasteur, Paris, France

[2]Unité Pasteur-CNAM Risques Infectieux et Emergents (PACRI), Conservatoire National des Arts et Métiers, Paris, France

[3]Service de Maladies Infectieuses et Tropicales, Hôpital Raymond Poincaré, Assistance Publique - Hôpitaux de Paris, Garches, France

[4]Service de Maladies Infectieuses et Réanimation Médicale, Centre Hospitalier Régional et Universitaire Pontchaillou, Rennes, France

[5]Infectious and Tropical Diseases Department, Bichat-Claude Bernard University Hospital and University of Paris, Assistance Publique - Hôpitaux de Paris, Paris, France

[6]Unité, Institut National de la Santé et de la Recherche Médicale, Paris, France

[7]Institut Inter-régional de Cancérologie Jean Bernard, Le Mans, France

Corresponding Author:
Fabrice Denis, MD, PhD
Institut Inter-régional de Cancérologie Jean Bernard
9 Rue Beauverger
Le Mans, 72100
France
Phone: 33 243475810
Email: f.denis@cjb72.org

Abstract

Background: We developed a self-triage web application for COVID-19 symptoms, which was launched in France in March 2020, when French health authorities recommended all patients with suspected COVID-19 call an emergency phone number.

Objective: Our objective was to determine if a self-triage tool could reduce the burden on emergency call centers and help predict increasing burden on hospitals.

Methods: Users were asked questions about their underlying conditions, sociodemographic status, postal code, and main COVID-19 symptoms. Participants were advised to call an emergency call center if they reported dyspnea or complete loss of appetite for over 24 hours. Data on COVID-19–related calls were collected from 6 emergency call centers and data on COVID-19 hospitalizations were collected from Santé Publique France and the French Ministry of Health. We examined the change in the number of emergency calls before and after the launch of the web application.

Results: From March 17 to April 2, 2020, 735,419 questionnaires were registered in the study area. Of these, 121,370 (16.5%) led to a recommendation to call an emergency center. The peak number of overall questionnaires and of questionnaires leading to a recommendation to call an emergency center were observed on March 22, 2020. In the 17 days preceding the launch of the web application, emergency call centers in the study area registered 66,925 COVID-19–related calls and local hospitals admitted 639 patients for COVID-19; the ratio of emergency calls to hospitalizations for COVID-19 was 104.7 to 1. In the 17 days following the launch of the web application, there were 82,347 emergency calls and 6009 new hospitalizations for COVID-19, a ratio of 13.7 calls to 1 hospitalization (chi-square test: $P<.001$).

Conclusions: The self-triage web application launch was followed by a nearly 10-fold increase in COVID-19–related hospitalizations with only a 23% increase in emergency calls. The peak of questionnaire completions preceded the peak of COVID-19–related hospitalizations by 5 days. Although the design of this study does not allow us to conclude that the self-triage tool alone contributed to the alleviation of calls to the emergency call centers, it does suggest that it played a role, and may be used for predicting increasing burden on hospitals.

Trial Registration: ClinicalTrials.gov NCT04331171; https://clinicaltrials.gov/ct2/show/NCT04331171

Introduction

Since February 2020, France has been hit by a severe COVID-19 epidemic that partly overwhelmed health system capacities. At the beginning of the epidemic, the French Ministry of Health recommended all patients with suspected COVID-19 call an emergency call center (Centre 15). As a result, patients experienced long delays before reaching an operator; some of these patients had a condition requiring emergency care. In this context, triage tools preselecting patients who should call the emergency call center may be particularly helpful. Web-based self-triage of symptoms is a growing field and has been shown to improve survival in oncology [1,2]. Past data have shown the feasibility of self-triage by parents of children with influenza-like illnesses, although specificity was weak [3]. Self-triage symptom checkers have higher levels of appropriate triage when used for emergency care, according to a study on symptom checkers that are available in Australia [4]. The use of web-based tools for COVID-19 management is currently increasing [5], but little data are available on self-triage and its impact on health care use. We sought to develop a web-based self-triage tool to optimize triage of patients with COVID-19 in France. A web application [6] was developed and launched while the COVID-19 epidemic was growing in France in March 2020. Our objective was to determine if a self-triage tool for COVID-19 could reduce the burden on emergency call centers and help predict increasing burden on hospitals.

Methods

The web application was launched on March 17, 2020, via a national media campaign in France including social media, radio, and magazine media. At that time, the French Ministry of Health recommended that all patients with suspected COVID-19 call an emergency call center. The recruitment process via the web application has already been detailed in a previous work [7]. Participants were asked about their postal code, pre-existing conditions, and potential COVID-19 symptoms (fever defined as body temperature >37.7 °C, unusual cough, shortness of breath, sore throat, muscle aches, diarrhea, loss of appetite, fatigue, anosmia, and ageusia). Depending on reported symptoms and underlying conditions, the user was recommended either to stay home and reuse the application in case of evolving symptoms (self-monitoring), or to contact a general practitioner (GP), or to call an emergency call center (if they reported shortness of breath or complete loss of appetite for over 24 hours) [8]. The web application did not offer monitoring of participant adherence to the self-triage recommendation. Access to the web application did not require a login or account creation. The web application did not identify participants who responded several times and did not follow up on participants. Questionnaires were excluded from the analysis if they did not include a postal code or if the completion duration

was considered inconsistent (<30 seconds). This study was approved by the French National Health-Data Institute, which reviews ethical conduct of human subject research, data confidentiality, and safety.

We collected data on COVID-19–related calls from 6 emergency call centers that cover some of the most severely COVID-19–affected areas in France (Bas-Rhin, Paris, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne), where burden was expected to be highest on emergency call centers, and one area where the web application was advertised through local papers a few days before the nationwide campaign, allowing for an earlier evaluation of impact (Sarthe). Data included calls made before the web application launch, from the day the first COVID-19–related hospitalization following an emergency room (ER) consultation was reported in the study area. That period, starting February 29, 2020, covers the 17 days preceding the launch of the web application. All of the territories covered in the study area had reported their first hospitalization following ER consultation for COVID-19 by March 3, 2020. We collected the same data the 17 days following the launch of the web application. Data regarding daily hospitalizations for COVID-19 following evaluation at an ER in the study area were provided by Santé Publique France and the French Ministry of Health. We compared the ratio of daily emergency center calls reported by emergency call centers to daily hospitalizations for COVID-19 before and after the launch of the web application using a chi-square test.

Results

From March 17 to April 2, 2020, there were 4,391,786 questionnaires filled out nationwide (Figure 1). Of these, 897,099 questionnaires were excluded from analysis for not including a postal code or inconsistent completion duration. The number of assessed questionnaires represent the number of assessments and not individuals. Among the 3,494,687 assessed questionnaires, 558,236 (16.0%) led to a recommendation to call an emergency call center. In the study area, 735,419 questionnaires were assessed, among which 121,370 (16.5%) led to a recommendation to call an emergency call center. Both the peak of overall questionnaires and that of questionnaires leading to a recommendation to call an emergency center were observed on March 22, 2020 (155,415 and 23,952, respectively; Figure 2).

The first hospitalization for COVID-19 following an ER consultation in the study area was reported on February 29, 2020. The peak of hospitalizations was observed on March 27, with 553 hospitalizations (Figure 2). In the 17 days preceding the launch of the web application, emergency call centers in the study area registered 66,925 COVID-19–related calls and local hospitals admitted 639 patients for COVID-19, a ratio of 104.7 calls to 1 hospitalization. In the 17 days following the launch

of the application, there were 82,347 COVID-19–related emergency calls (a 23% increase from the previous period) and 6009 new hospitalizations for COVID-19 (a 9.4-fold increase from the previous period), resulting in a ratio of 13.7 calls to 1 hospitalization (chi-square test: *P*<.001; Figure 2).

Figure 1. Flowchart of self-triage web application respondents.
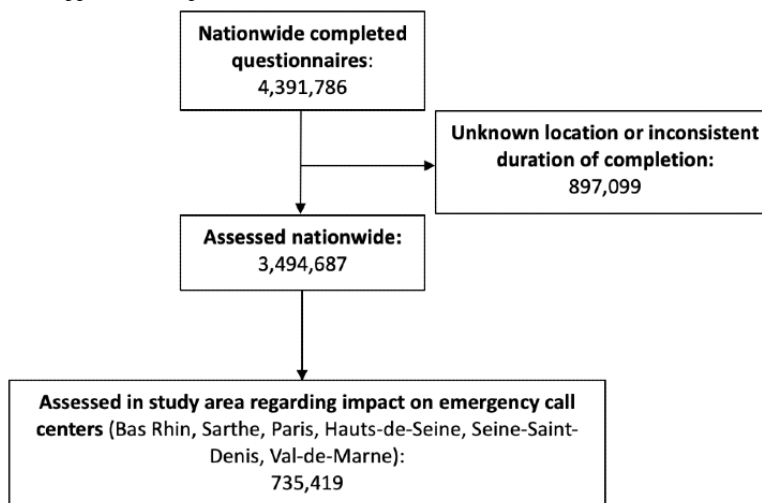
Figure 2. Data from the study area (raw numbers). (A) COVID-19–related calls to emergency call centers and hospitalizations following a COVID-19–related emergency room consultation. (B) Overall assessed questionnaires and the number of recommendations to call an emergency center. The web application was launched March 17, 2020 (dashed line).
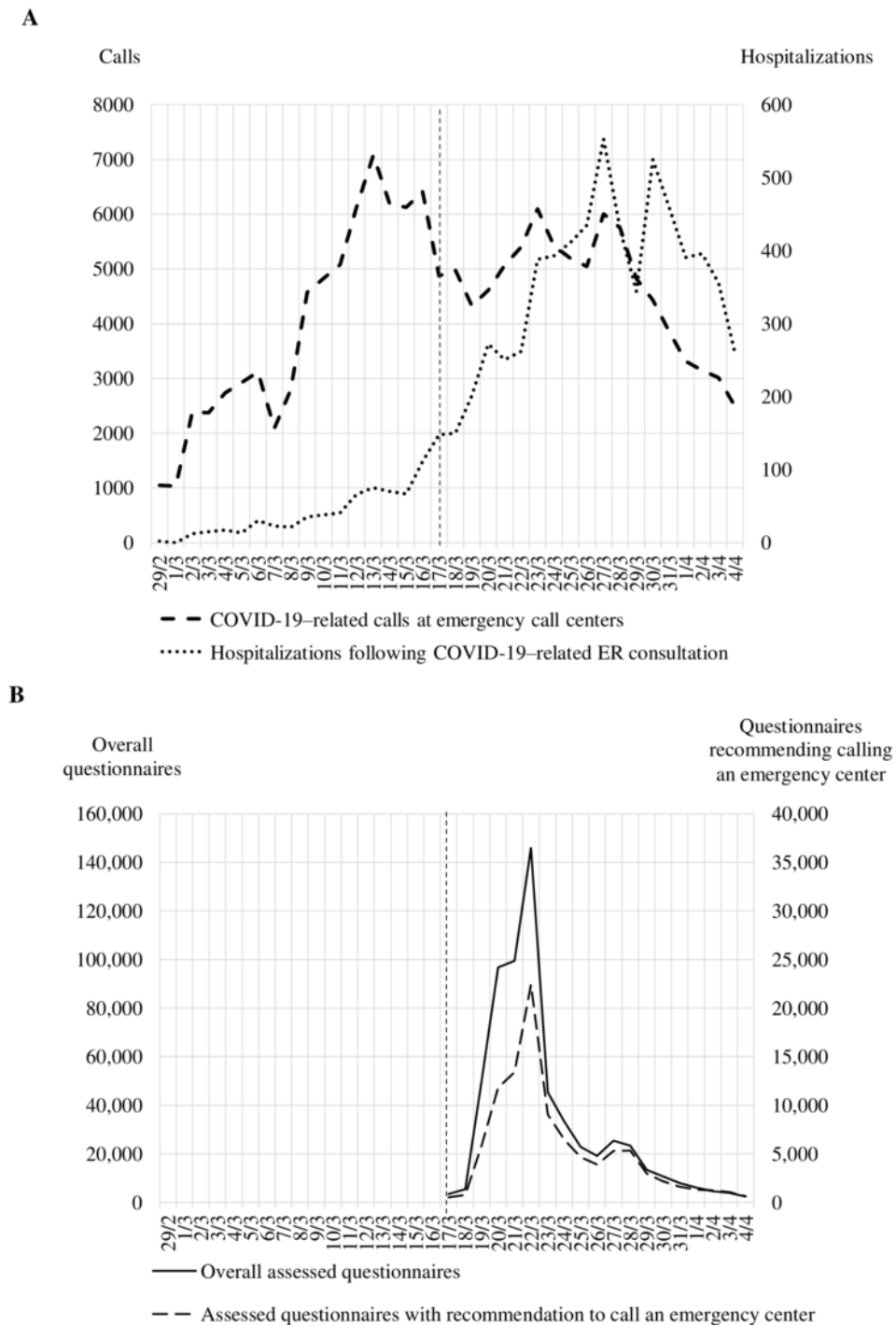
Discussion

The launch of the self-triage web application was followed by a nearly 10-fold increase in COVID-19–related hospitalizations with only a 23% increase in emergency calls, even though the number of completed questionnaires quickly surged, including questionnaires leading to a recommendation to call an emergency center, indicating appropriate use of the tool. Both the peak of overall questionnaires and that of questionnaires leading to a recommendation to call an emergency center happened 5 days after lockdown started in France on March 17, 2020. This is compatible with a maximum incidence rate of SARS-CoV-2 infections one day before lockdown, considering a mean 5-day incubation period for COVID-19 [9,10]. There

was a further delay of 5 days between the peak of questionnaire completions and the peak of COVID-19–related hospitalizations, consistent with a mean duration between infection and hospitalization for severe forms of the disease of 10 days, as previously described [11]. It is unknown how many questionnaires were filled out by people with COVID-19. However, the nationwide daily incidence immediately before the lockdown onset on March 17, 2020, was estimated to be between 180,000 and 490,000 in a study by Salje et al [12]. The positive predictive value of general symptoms (eg, dyspnea or loss of appetite) increases in such a high-incidence setting, suggesting a significant share of people reporting symptoms during the surge peaking on March 22 had COVID-19, although more precise evaluation is impossible. It indicates that the self-triage tool could help predict a rise in severe cases and burden on hospitals. This hypothesis needs confirmation should a new surge in COVID-19 cases and related hospitalizations occur.

There are few data regarding the impact of self-triage tools on health care use. A recent systematic review of self-triage symptom checkers for urgent health problems suggested they led to less frequent health care use [13]. Verzantvoort et al [14] reported 67% of patients receiving self-care advice intended to follow the advice. For COVID-19, Judson et al [15] described a dedicated self-triage tool that recommended self-care to 40% of symptomatic patients; that advice was mostly followed, as only 8% of them had an in-person visit in the following 48 hours, suggesting an effective reduction in unnecessary GP or ER visits [15].

The design of the present study does not allow us to conclude that the self-triage tool alone contributed to the alleviation of calls to the emergency call centers. Other interventions, such as the creation of an information hotline for nonurgent COVID-19–related questions, happened soon after the launch of the web application and may have contributed to relieving the burden on emergency call centers. The sudden drop in completed questionnaires following the peak may be related to the drop in infections after lockdown, but may also indicate the influence of media campaigns that promoted the use of the web application. Interpretation of peaks in website usage should therefore be cautious and take into account those campaigns. Altogether, we think that a self-triage tool can be useful in periods of high incidence of COVID-19, when health care use quickly surges and health care providers such as emergency call centers endure a rapidly increasing burden. Helping predict increasing burden on hospitals may also help policy makers and health care providers by informing their response.

Conflicts of Interest

None declared.

References

1. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]

2. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]

3. Anhang Price R, Fagbuyi D, Harris R, Hanfling D, Place F, Taylor TB, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: a cautionary tale. JAMA Pediatr 2013 Feb;167(2):112-118. [doi: 10.1001/jamapediatrics.2013.1573] [Medline: 23254373]

4. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]

5. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. JMIR mHealth uHealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]

6. Maladie Coronavirus. URL: https://www.maladiecoronavirus.fr [accessed 2020-06-09]

7. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological Observations on the Association Between Anosmia and COVID-19 Infection: Analysis of Data From a Self-Assessment Web Application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]

8. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]

9. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]

10. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine 2020 May 05;172(9):577-582. [doi: 10.7326/m20-0504]

11. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. In: The Lancet Infectious Diseases. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis Internet Elsevier; Aug 2020:911-919.

12. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. Science 2020 Jul 10;369(6500):208-211 [FREE Full text] [doi: 10.1126/science.abc3517] [Medline: 32404476]

13. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]

14. Verzantvoort NCM, Teunis T, Verheij TJM, van der Velden AW. Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? PLoS One 2018;13(6):e0199284 [FREE Full text] [doi: 10.1371/journal.pone.0199284] [Medline: 29944708]

15. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

Abbreviations

ER: emergency room
GP: general practitioner

Original Paper

Impact of Trump's Promotion of Unproven COVID-19 Treatments and Subsequent Internet Trends: Observational Study

Kacper Niburski[1], BSc, MA; Oskar Niburski[2], BSc

[1]McGill University, Montreal, QC, Canada

[2]York University, Toronto, ON, Canada

Corresponding Author:
Kacper Niburski, BSc, MA
McGill University
3655 Promenade Sir William Osler
Montreal, QC, H3G 0B1
Canada
Phone: 1 9055162020
Email: kacperniburski@gmail.com

Abstract

Background: Individuals with large followings can influence public opinions and behaviors, especially during a pandemic. In the early days of the pandemic, US president Donald J Trump has endorsed the use of unproven therapies. Subsequently, a death attributed to the wrongful ingestion of a chloroquine-containing compound occurred.

Objective: We investigated Donald J Trump's speeches and Twitter posts, as well as Google searches and Amazon purchases, and television airtime for mentions of hydroxychloroquine, chloroquine, azithromycin, and remdesivir.

Methods: Twitter sourcing was catalogued with Factba.se, and analytics data, both past and present, were analyzed with Tweet Binder to assess average analytics data on key metrics. Donald J Trump's time spent discussing unverified treatments on the United States' 5 largest TV stations was catalogued with the Global Database of Events, Language, and Tone, and his speech transcripts were obtained from White House briefings. Google searches and shopping trends were analyzed with Google Trends. Amazon purchases were assessed using Helium 10 software.

Results: From March 1 to April 30, 2020, Donald J Trump made 11 tweets about unproven therapies and mentioned these therapies 65 times in White House briefings, especially touting hydroxychloroquine and chloroquine. These tweets had an impression reach of 300% above Donald J Trump's average. Following these tweets, at least 2% of airtime on conservative networks for treatment modalities like azithromycin and continuous mentions of such treatments were observed on stations like Fox News. Google searches and purchases increased following his first press conference on March 19, 2020, and increased again following his tweets on March 21, 2020. The same is true for medications on Amazon, with purchases for medicine substitutes, such as hydroxychloroquine, increasing by 200%.

Conclusions: Individuals in positions of power can sway public purchasing, resulting in undesired effects when the individuals' claims are unverified. Public health officials must work to dissuade the use of unproven treatments for COVID-19.

Introduction

Numerous treatments have been suggested for the novel COVID-19 disease, with only remdesivir recently showing some potential efficacy prior to its approval on May 1, 2020 [1]. The efficacy of other therapies, such as hydroxychloroquine, chloroquine, and azithromycin, remains unproven, though they have been praised by numerous public figures, such as US president Donald J Trump.

Previous studies have shown that individuals with high influence (ie, individuals with large social capital or political power) [2] can affect public decisions and purchasing power [3]. Moreover, many of these individuals have the ability to affect content, sentiment, and public attention by using social media to

disseminate information [4]. Donald J Trump is one such individual who has previously influenced behavior through his high social presence on the internet and his political capital as the president of the United States [5].

In 2020, during the initial period of the COVID-19 pandemic, many people were focused on uninvestigated therapies, especially Donald J Trump. However, it has become increasingly clear that some therapies, such as hydroxychloroquine, had numerous concerns surrounding them. Studies have reported that hydroxychloroquine can cause conduction disturbances and fatal arrhythmias [6], the supply of hydroxychloroquine may decrease for approved conditions like rheumatoid arthritis [7], and chloroquine is similar to chloroquine products that are used as aquarium cleaners, which may be toxic if ingested in large quantities [8]. A death has already occurred due to the ingestion of chloroquine products [9].

A recent study has noted that there was an influx of internet searches for hydroxychloroquine and chloroquine after being tweeted about by US president Donald J Trump [10]. However, the study did not investigate all therapies that were initially suggested in the United States, the purchasing amounts on Google and Amazon, or the television airtime for each respective medication. The study also did not associate these purchasing amounts with Donald J Trump's discussions and tweets, which are 2 avenues where his reach is in the millions. This would have better shown a relationship between his influence and the impact he has on his supporters and the general public.

We therefore investigated the relationship between Donald J Trump's advocacy for unproven treatments and the media landscape of COVID-19 treatments. We also investigated how these coalesced into behavioral changes in individuals. We analyzed the time Donald J Trump spent discussing hydroxychloroquine, chloroquine, azithromycin, and remdesivir on the 5 largest US television stations, the total number of searches and purchases on Google and Amazon for these treatments, and the Twitter statistics and fallout of Donald J Trump's advocacy for these therapies.

Methods

Twitter

Tweets were catalogued based on mentions of hydroxychloroquine, chloroquine, azithromycin, and remdesivir. Factba.se was used to note if any Twitter posts were archived or otherwise hidden by Donald J Trump [11]. Tweet Binder was used to assess average analytics data on key metrics, such as likes, retweets, and comments [12]. Tweet Binder further aggregated all tweets and averaged longitudinal tweeting patterns.

Television

The time Donald J Trump spent discussing each treatment on television was recorded from the 5 largest US television stations: CNN (Cable News Network), C-SPAN (Cable-Satellite Public Affairs Network), Fox News, MSNBC (Microsoft/National Broadcasting Company), and BBC (British Broadcasting Corporation) News. The Global Database of Events, Language, and Tone was used to note White House briefings and the total

monitored airtime of each station for broadcasts that aired from March 1 to April 30, 2020 [13]. Airtime data are presented as percentages of 15-second airtime blocks, which is the maximum the software can crawl its historical database. Manual crawling of television airtimes or news headings was not performed.

Google

Google searches from March 1 to April 30, 2020 were indexed with Google Trends, focusing on searches performed within the United States [14]. Searches including the words "hydroxychloroquine," "chloroquine," "azithromycin," "remdesivir," and "covid treatment" were catalogued, along with the Google purchasing patterns for these items. Analyses for these searches were performed concurrently with cataloguing. However, the results for this period were not large enough to report. The data obtained were the proportional patterns of searches instead of exact amounts, given Google's data agreement [15].

Amazon

Amazon purchases were indexed based on the following words: "hydroxychloroquine," "chloroquine," "azithromycin," and "remdesivir." Search volume was assessed using Helium 10 software [16]. Estimated purchase amounts were calculated using Ahrefs [17].

Data Analysis

Data were analyzed with Microsoft Excel. This included aggregating the data, cataloguing data, and formatting data with key points, such as Donald J Trump's first press conference on March 19, 2020 or his numerous tweeting dates.

The wide variety of sampling, numerous software-defined metrics, and methods of measuring outcomes, such as gross estimates and proportions instead of true samples, did not lend themselves to statistical analysis [18]. The key assumption of transitivity, which assumes no important differences in the distribution of either potential effect modifiers or sampling techniques (eg, individual comparisons of price points between Google and Amazon), would be high for indirect comparisons.

Results

Twitter

Table 1 notes the characteristics of US president Donald J Trump's tweets. Donald J Trump mainly focused on hydroxychloroquine, as per his own tweets (eg, his tweets on March 21, 2020) and his retweets of Twitter posts from other people who supported his claims regarding the efficacy of hydroxychloroquine. Various tweets included news articles (eg, the tweet on April 4, 2020), which were mostly from Fox News (4/6, 67%).

Based on our analysis with Tweet Binder [19], with an average proportion of nearly 100,000 likes and 20,000 retweets, Donald J Trump's first tweet on March 21, 2020 advocating for the use of hydroxychloroquine and azithromycin was one of his most popular tweets. With 385,700 likes and 103,200 retweets at the time of this study, the tweet had a potential estimated impression

reach (ie, the number of potential user views on the individual   tweet) of 78,800,580.

Table 1. Characteristics Donald J Trump's tweets that mention hydroxychloroquine, chloroquine, chloroquine, and remdesivir, including key metrics such as retweets, likes, and whether the tweets were retweets or self-composed tweets by Donald J Trump.

Therapy/Mention	Date	Tweet sources	Number of retweets of other posts	Number of self-retweets	Number of retweets	Number of likes
Hydroxychloroquine						
1	April 18, 2020	@paulsperry_ [20]	1	0	12,600	30,500
2	April 10, 2020	@cybergenica [21]	1	0	5800	16,900
3	April 4, 2020	@OANN [22]	1	0	7800	21,600
4	April 4, 2020	@jsolomonReports [23]	1	0	11,300	28,200
5	March 23, 2020	@AndrewCMcCarthy [24]	1	0	9100	23,700
6	March 21, 2020	@realDonaldTrump [25]	0	1	103,200	385,700
7	March 21, 2020	@MichaelCoudrey [26]	1	0	27,300	53,900
8	March 21, 2020	@realDonaldTrump [27]	1	0	103,200	385,700
Azithromycin						
1	March 21, 2020	@realDonaldTrump [25]	0	1	103,200	385,700
2	March 21, 2020	@MichaelCoudrey [26]	1	0	27,300	53,900
3	March 21, 2020	@realDonaldTrump [27]	0	0	103,200	385,700
Chloroquine						
0	N/A[a]	0	0	0	0	0
Remdesivir						
0	N/A	0	0	0	0	0

[a]N/A: not applicable.

Television Mentions and Broadcasts

Table 2 notes the number of times Donald J Trump mentioned COVID-19 treatments during White House briefings. He mentions hydroxychloroquine 37 times in total, most frequently on April 4, 2020 (9 times); chloroquine 12 times in total, most frequently on March 19, 2020 (7 times); azithromycin 8 times in total, most frequently on March 19, 2020 (3 times, while also mentioning hydroxychloroquine); and remdesivir 8 times in total, most frequently on March 19, 2020 (4 times).

Table 2. Dates of Donald J Trump's mentions of unproven therapies during televised appearances.

Date	Number of mentions on television			
	Hydroxychloroquine	Chloroquine	Azithromycin	Remdesivir
March 19, 2020	4	7	3	4
March 21, 2020	3	0	0	1
March 23, 2020	2	0	0	0
March 27, 2020	1	1	1	0
March 29, 2020	1	0	0	0
March 31, 2020	4	2	0	0
April 1, 2020	1	1	0	1
April 4, 2020	9	0	0	0
April 5, 2020	3	0	0	1
April 6, 2020	1	0	0	0
April 7, 2020	3	0	1	0
April 8, 2020	1	0	0	0
April 9, 2020	1	0	1	0
April 10, 2020	2	0	1	0
April 13, 2020	1	1	1	1
Total	37	12	8	8

Figure 1 displays the television airtime of broadcasts that mentioned the treatments. Airtime for mentions of chloroquine (Figure 1A) peaked on March 24, 2020 on Fox News (0.59%), MSNBC (0.78%), and CNN (0.61%), following the chloroquine-related death. Following Donald J Trump's tweet advocating for the use of hydroxychloroquine, airtime for mentions of hydroxychloroquine (Figure 1B) peaked on April 4 and April 6, 2020 on Fox News (1.94% and 2.46%, respectively), while peaking on CNN on April 22, 2020 (2.88%). Fox News was one of the single television networks that continually provided coverage on azithromycin (Figure 1C), as airtime for its mention peaked at 0.45% on April 7, 2020. Remdesivir (Figure 1D) was not mentioned on any news organizations until April 30, 2020, followed by the decision by the US Food and Drug Administration to approve remdesivir as a COVID-19 treatment on May 1, 2020 [28]. At this point, airtime for mentions of remdesivir experienced peaks comparable to those of the other treatments. COVID-19 treatment coverage has been increasing on all networks, especially after the March 19, 2020 press conference.

XSL•FO
RenderX

Figure 1. Television airtime of keywords from March 1 to April 30, 2020 during 15-second airtime blocks. Dark blue represents BBC News, orange represents CNN, grey represents C-SPAN, yellow represents Fox News, and light blue represents MSNBC. (A) Airtime for the term "chloroquine". (B) Airtime for the term "hydroxychloroquine." (C) Airtime for the term "azithromycin." (D) Airtime for the term "remdesivir". (E) Airtime for the term "covid treatment". CNN: Cable News Network; C-SPAN: Cable-Satellite Public Affairs Network.
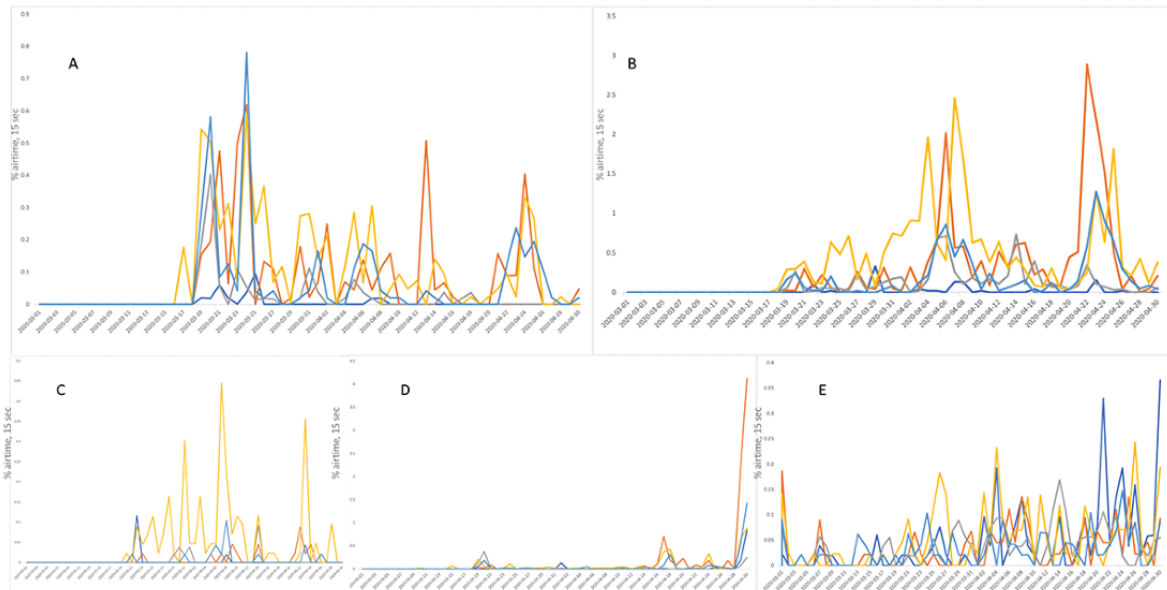
Google and Amazon

Searches for and purchases (Figure 2A) of chloroquine on Google peaked on March 19, 2020, followed by a second peak on March 23, 2020, after the story of chloroquine-poisoning was made public. Searches for and purchases of hydroxychloroquine followed a similar trend, though there was a second peak on April 4, 2020 when Donald J Trump retweeted a studying discussing the apparent efficacy of hydroxychloroquine. Peaks in COVID-19 treatment searches (Figure 1E) largely coincided with Donald J Trump's White House briefing on March 19, 2020, with 3 peaks correlating with his tweet history in Table 1 (March 19, March 21, and April 4, 2020).

Figure 2. Google searches (blue line) for and purchases (orange line) of keywords. (A) Results for the term "chloroquine." (B) Results for the term "hydroxychloroquine." (C) Results for the term "azithromycin." (D) Results for the term "remdesivir." (E) Results for the term "covid treatment.".
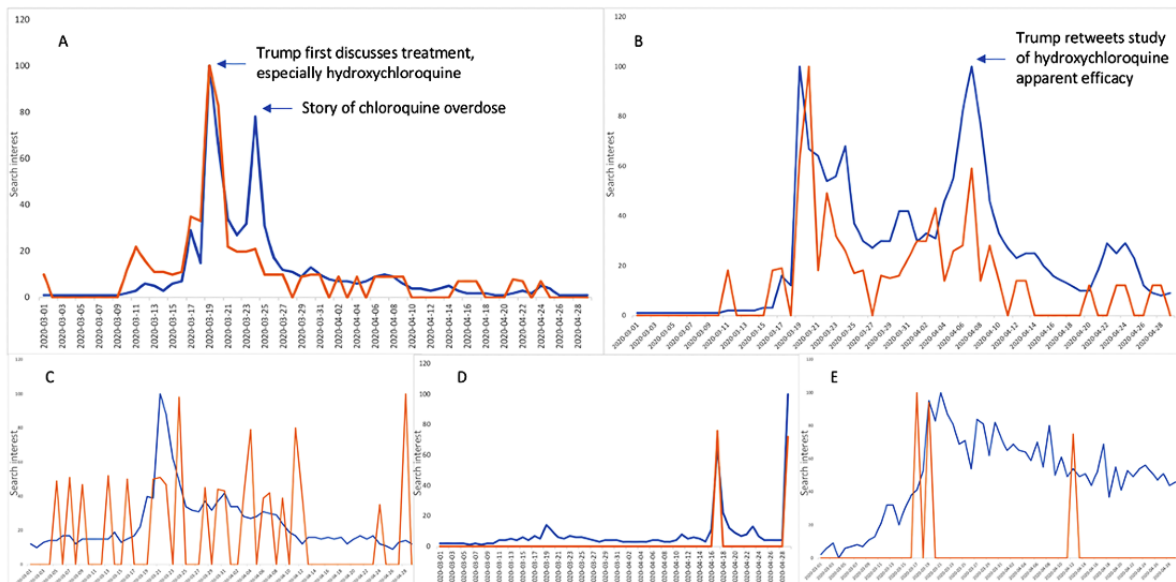
Our Amazon purchase analysis (Figure 3) showed a peak in purchases for chloroquine, hydroxychloroquine, and azithromycin after the March 19, 2020 White House briefing. This increase in purchases was observed for various forms of these medications, whether it was for purchases of a book about azithromycin [29], which increased by 10 sales; herbal elements claiming to have hydroxychloroquine as an active element [30], which increased by an estimated 50 sales; or texts on alternative chloroquine-containing compounds, such as chloroquine phosphate, which increased by an estimated 30 sales [31]. Secondary peaks for searches on and purchases for hydroxychloroquine were observed after the tweet on April 4, 2020.

Figure 3. Amazon purchases of alternate therapies. Grey represents purchases of hydroxychloroquine, blue represents purchases of azithromycin, and orange represents purchases of chloroquine.
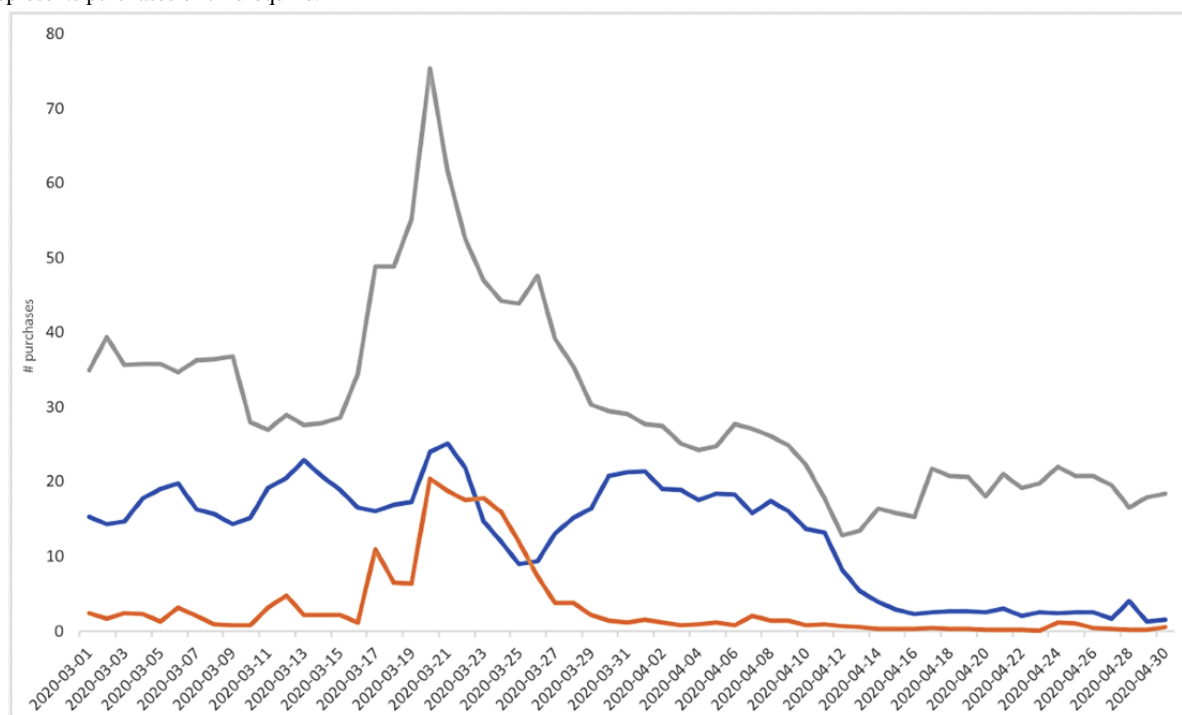
Discussion

This study is the first to show the media landscape of COVID-19 therapies prior to gaining approval or complete disregard from the scientific community. Our results show that there was a substantial increase in purchases and searches for previously unpurchased and unsearched therapies by the general public following the backing of US president Donald J Trump. These increases correlated with his discussions in press conferences and personal social media posts advocating for hydroxychloroquine and chloroquine cures. Conservative outlets provided the most airtime to hydroxychloroquine and chloroquine, and airtime for these therapies only peaked on liberal media outlets after the chloroquine-induced death occurred. Many of the increases in the purchases of these products mirror the increases in searches and airtimes that followed the initial press conference on March 19, 2020, during which Donald J Trump touted these therapies, and the subsequent social media endorsements on Twitter by Donald J Trump. Remdesivir, the treatment with the most evidence from small preliminary studies and largescale testing for COVID-19, had the least coverage time from all stations assessed in this study until just before its emergency approval on May 1, 2020. Donald J Trump has not advocated for its use publicly, except during the initial press conference on March 19, 2020, when all the suggested treatments were outlined.

During unknown medical situations, delicacy is required when determining the best treatments. Previous studies have shown that individuals are susceptible to easy claims and conspiracies without appropriate evidence [32], and once these inauthentic claims are given momentum, they are hard to dissuade [33]. It is for this reason that individuals often seek out influential figures for guidance and knowledge [34]. Providing assurances for unverified claims and treatments is dangerous, given that medications can have numerous side effect profiles. Recent trials have in fact been halted due to their general risk, as was the case with hydroxychloroquine [35]. Additionally, their utility for approved conditions, such as malaria, has been compromised due to limited access and hoarding [36].

Efforts need to be made to prevent further harm. In some instances, this has already occurred. Google has decreased access to links that sell chloroquine, whereas Amazon has removed links to chloroquine phosphate and provided COVID-19–related information on associated links. However, this is only true for the US version of the website. Other domains, such as .ca, still provide easy purchasing for medicine substitutes [37]. Public heath individuals must do more to advocate for safer, evidenced-based approaches. For example, Twitter has recently provided users the ability to flag COVID-19 misinformation and take down posts. This has occurred to Donald J Trump twice after April 30, 2020 for making nonfactual hydroxychloroquine claims [5].

There are limitations in this study. The estimated number of purchases and searches on Google and Amazon were both estimates, as they were provided by external providers rather than the services themselves. However, given the limited data access, these estimates are the best available to third party providers and researchers. Still, ours is first study to examine the purchasing behaviors of individuals and the media landscape of COVID-19 treatments following Donald J Trump's endorsements. Future studies will examine the searches and purchases of certain key words/items, such as masks, ultraviolet, and disinfectants.

Conflicts of Interest

None declared.

References

1. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet 2020 May 16;395(10236):1569-1578 [FREE Full text] [doi: 10.1016/S0140-6736(20)31022-9] [Medline: 32423584]
2. Schoorman FD, Mayer R, Davis JH. Social influence, social interaction, and social psychology in the study of trust. Acad Manage Rev 1996;21(2):337-339.
3. Cao P, Meister S, Klante O. How Social Media Influence Apparel Purchasing Behavior. Marketing Review St. Gallen 2014 Dec 9;31:77-86 [FREE Full text] [doi: 10.1365/s11621-014-0427-y]
4. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
5. Silverstein J. Twitter flags Trump tweet with fact-checking label for first time. CBS News. 2020 May 27. URL: https://www.cbsnews.com/news/twitter-adds-fact-check-warning-trump-tweets/ [accessed 2020-11-11]
6. Roos JM, Aubry M, Edwards WD. Chloroquine cardiotoxicity: clinicopathologic features in three patients and comparison with three patients with Fabry disease. Cardiovasc Pathol 2002;11(5):277-283. [doi: 10.1016/s1054-8807(02)00118-7] [Medline: 12361838]
7. Morris SJ, Wasko MCM, Antohe JL, Sartorius JA, Kirchner HL, Dancea S, et al. Hydroxychloroquine use associated with improvement in lipid profiles in rheumatoid arthritis patients. Arthritis Care Res (Hoboken) 2011 Apr;63(4):530-534 [FREE Full text] [doi: 10.1002/acr.20393] [Medline: 21452265]
8. Mégarbane B. Chloroquine and hydroxychloroquine to treat COVID-19: between hope and caution. Clin Toxicol (Phila) 2020 Apr 02:1-2. [doi: 10.1080/15563650.2020.1748194] [Medline: 32237918]
9. Theresa W, Alsup D, McLaughlin EC. Fearing coronavirus, Arizona man dies after taking a form of chloroquine used to treat aquariums. CNN. 2020 Mar 25. URL: https://www.cnn.com/2020/03/23/health/arizona-coronavirus-chloroquine-death/index.html [accessed 2020-05-06]
10. Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet Searches for Unproven COVID-19 Therapies in the United States. JAMA Intern Med 2020 Aug 01;180(8):1116-1118 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1764] [Medline: 32347895]
11. Donald Trump Complete - Best Tweets, Speeches, Policies. Factbase. URL: https://factba.se/ [accessed 2020-05-06]
12. The Most Complete Hashtag Tracking Tool For Twitter. Tweet Binder. URL: https://www.tweetbinder.com/ [accessed 2020-05-06]
13. Television Explorer. GDELT Summary. URL: https://api.gdeltproject.org/api/v2/summary/summary?d=iatv&t=summary [accessed 2020-05-06]
14. Explore what the world is searching. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2020-05-06]
15. Google. URL: https://policies.google.com/privacy?hl=en-US [accessed 2020-11-11]
16. Everything You Need to Sell on Amazon, and More. Helium 10. URL: https://www.helium10.com/ [accessed 2020-05-06]
17. Ahrefs. Keywords Explorer by Ahrefs: Discover Keyword Ideas and Analyze SEO Metrics. URL: https://ahrefs.com/keywords-explorer [accessed 2020-05-06]
18. Cumming G. Understanding The New Statistics Effect Sizes, Confidence Intervals, and Meta-Analysis. New York: Routledge; 2012.
19. Donald Trump on Twitter – 2009 / 2020 analysis. Tweet Binder. URL: https://www.tweetbinder.com/blog/trump-twitter/ [accessed 2020-05-06]
20. @paulsperry_. Twitter. 2020 Apr 13. URL: https://twitter.com/realdonaldtrump/status/1251587393955729412 [accessed 2020-11-13]
21. @cybergenica. Macron faces calls to allow wider use hydroxychloroquine in coronavirus fight: report https://ift.tt/2XrLh6i #FoxNews #CNN #MSNBC. Twitter. 2020 Apr 10. URL: https://twitter.com/realdonaldtrump/status/1248586394311761920 [accessed 2020-11-13]
22. @OANN. Hydroxychloroquine proving an effective treatment for coronavirus patien... https://youtu.be/NyDjJ5ytc54 #OANN @PearsonSharp. Twitter. 2020 Apr 01. URL: https://twitter.com/realdonaldtrump/status/1246627587289382918 [accessed 2020-11-13]
23. @jsolomonReports. Hydroxychloroquine rated top drug by doctors on front line of COVID-19 who also fear a second wave. Twitter. 2020 Apr 03. URL: https://twitter.com/realdonaldtrump/status/1246458783548702726 [accessed 2020-11-13]
24. @AndrewCMcCarthy. Our experience suggests that hydroxychloroquine should be a first-line treatment for Covid-19. We can use it to save lives and prevent others from becoming infected, write @DrJeffColyer and Daniel Hinthorn. 2020 Mar 23. URL: https://twitter.com/realdonaldtrump/status/1242302644498108416 [accessed 2020-11-13]
25. @realDonaldTrump. Twitter. 2020 Mar 21. URL: https://twitter.com/realdonaldtrump/status/1241579259249991684 [accessed 2020-11-13]

26.   @MichaelCoudrey. Twitter. 2020 Mar 21. URL: https://twitter.com/realdonaldtrump/status/1241398266941657093 [accessed 2020-11-13]

27.   @realDonaldTrump. HYDROXYCHLOROQUINE & AZITHROMYCIN, taken together, have a real chance to be one of the biggest game changers in the history of medicine. The FDA has moved mountains - Thank You! Hopefully they will BOTH (H works better with A, International Journal of Antimicrobial Agents). 2020 Mar 21. URL: https://twitter.com/realdonaldtrump/status/1241367239900778501 [accessed 2020-11-13]

28.   Coronavirus (COVID-19) Update: FDA Issues Emergency Use Authorization for Potential COVID-19 Treatment. U.S. Food & Drug Administration. URL: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-emergency-use-authorization-potential-covid-19-treatment [accessed 2020-05-06]

29.   Taste Masked Azithromycin Dispersible Tablets for Pediatrics: Pharmaceutical Technology Paperback – July 31, 2012. Amazon. URL: https://www.amazon.com/dp/3659189243 [accessed 2020-05-06]

30.   Chloroxygen Chlorophyll Concentrate: 120-Count Softgels. Amazon. URL: https://www.amazon.com/dp/B00HXQZRA2 [accessed 2020-05-06]

31.   Chloroquine Phosphate; Second Edition 2nd Edition. Amazon. URL: https://www.amazon.com/dp/1984235532 [accessed 2020-05-06]

32.   Lobato E, Mendoza J, Sims V, Chin M. Examining the Relationship Between Conspiracy Theories, Paranormal Beliefs, and Pseudoscience Acceptance Among a University Population. Appl Cognit Psychol 2014 Jun 02;28(5):617-625. [doi: 10.1002/acp.3042]

33.   Bishop FL, Jacobson EE, Shaw JR, Kaptchuk TJ. Scientific tools, fake treatments, or triggers for psychological healing: how clinical trial participants conceptualise placebos. Soc Sci Med 2012 Mar;74(5):767-774 [FREE Full text] [doi: 10.1016/j.socscimed.2011.11.020] [Medline: 22285289]

34.   Lipsman A, Mudd G, Rich M, Bruich S. The Power of "Like". J Advert Res 2012 Mar 01;52(1):40-52. [doi: 10.2501/jar-52-1-040-052]

35.   Touret F, de Lamballerie X. Of chloroquine and COVID-19. Antiviral Res 2020 May;177:104762 [FREE Full text] [doi: 10.1016/j.antiviral.2020.104762] [Medline: 32147496]

36.   Martins-Filho P, Carvalho AC, de Melo EM, Mendes MLT, Santos V. The "unbridled race" for using chloroquine and hydroxychloroquine to prevent or treat COVID-19 leads to shortages for patients with chronic inflammatory conditions and malaria in Brazil. CMAJ. URL: https://www.cmaj.ca/content/unbridled-race-using-chloroquine-and-hydroxychloroquine-prevent-or-treat-covid-19-leads [accessed 2020-05-06]

37.   Chloroquine phosphate: Pet Supplies. Amazon. URL: https://www.amazon.ca/s?k=chloroquine+phosphate&i=pets&dc&qid=1588730174&ref=sr_nr_i_9 [accessed 2020-05-06]

Abbreviations

BBC: British Broadcasting Corporation
CNN: Cable News Network
C-SPAN: Cable-Satellite Public Affairs Network
MSNBC: Microsoft/National Broadcasting Company

XSL•FO

RenderX

Original Paper

COVID-19–Related Internet Search Patterns Among People in the United States: Exploratory Analysis

Tony S Shen[1], MD; Aaron Z Chen[2], BA; Patawut Bovonratwet[1], MD; Carol L Shen[3], MD; Edwin P Su[1], MD

[1]Hospital for Special Surgery, New York, NY, United States

[2]Weill Cornell Medical College, New York, NY, United States

[3]NewYork-Presbyterian Morgan Stanley Children's Hospital, New York, NY, United States

Corresponding Author:
Aaron Z Chen, BA
Weill Cornell Medical College
1300 York Avenue
New York, NY, 10065
United States
Phone: 1 8605196413
Email: azc2001@med.cornell.edu

Abstract

Background: The internet is a well-known source of information that patients use to better inform their opinions and to guide their conversations with physicians during clinic visits. The novelty of the recent COVID-19 outbreak has led patients to turn more frequently to the internet to gather more information and to alleviate their concerns about the virus.

Objective: The aims of the study were to (1) determine the most commonly searched phrases related to COVID-19 in the United States and (2) identify the sources of information for these web searches.

Methods: Search terms related to COVID-19 were entered into Google. Questions and websites from Google web search were extracted to a database using customized software. Each question was categorized into one of 6 topics: clinical signs and symptoms, treatment, transmission, cleaning methods, activity modification, and policy. Additionally, the websites were categorized according to source: World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), non-CDC government, academic, news, and other media.

Results: In total, 200 questions and websites were extracted. The most common question topic was transmission (n=63, 31.5%), followed by clinical signs and symptoms (n=54, 27.0%) and activity modification (n=31, 15.5%). Notably, the clinical signs and symptoms category captured questions about myths associated with the disease, such as whether consuming alcohol stops the coronavirus. The most common websites provided were maintained by the CDC, the WHO, and academic medical organizations. Collectively, these three sources accounted for 84.0% (n=168) of the websites in our sample.

Conclusions: In the United States, the most commonly searched topics related to COVID-19 were transmission, clinical signs and symptoms, and activity modification. Reassuringly, a sizable majority of internet sources provided were from major health organizations or from academic medical institutions.

Introduction

Since its emergence in late 2019 in Wuhan, China, COVID-19, the disease caused by the novel coronavirus SARS-CoV-2 has drastically impacted daily life around the world [1,2]. Among the changes to the public sphere include government-imposed lockdowns of businesses, schools, and universities, designed to mitigate the spread of the disease and to alleviate the significant strain on health care systems [2]. As many continue to stay at home due to the COVID-19 pandemic, internet use has become an increasingly dominant part of daily life. In a recent poll, a majority of Americans considered the internet "essential" during this time [3]. Nearly all major internet services have seen increased traffic since early March 2020 [4]. Given the

unprecedented nature of the pandemic, there is naturally much public uncertainty regarding COVID-19, and thus, many are turning to the internet to ask their questions and obtain information about the coronavirus.

Previous studies have shown that patients frequently use the internet to research their conditions and inform their discussions in clinic [5,6]. As physicians, insight into what topics patients are curious or anxious about may help guide and structure our interactions, leading to improved patient rapport and satisfaction. Additionally, with well-publicized recent examples of misinformation originating from many sources, including places of authority, it is paramount for physicians to collectively take responsibility to provide reliable and trustworthy information based on the best available evidence [7,8]. Thus, the aims of the present study were to (1) determine the most commonly searched phrases related to COVID-19 in the United States and (2) identify the sources of information for these web searches. In doing so, we believe that we can distill the collective curiosity of the internet-using public into useful information for physicians in clinic.

Methods

Search terms related to COVID-19 were entered into Google web search using a clean-installed Google Chrome browser on May 30, 2020, in New York, NY. Google web search is by far the most widely used internet search engine in the United States [9]. In 2018, Google introduced a natural language processing algorithm, which greatly improved the ability of the search engine to identify clusters of search queries related to any given topic [10]. Due to this new technology, Google redirects all searches related to COVID-19, such as "COVID-19," "coronavirus," "coronavirus disease," etc, to a centralized COVID-19 homepage. This search results page incorporates the location of the user's search and generates a list of questions and websites that are frequently associated with the initial query. On each results page, the 200 most commonly asked questions were generated. The questions were downloaded to a database using a freely available program (Scraper, version 1.7). The specific question and web address were noted on the webpage by their unique XML Path Language (XPath) strings.

The questions were first categorized according to Rothwell's classification system by a single trained reviewer [11,12]. This classification system was expanded into one of 6 topics—clinical signs and symptoms, treatment, transmission, cleaning methods, activity modification, and policy—which were chosen based on previously published studies that examined the web and social media concerns of users during the COVID-19 pandemic [13-15] (Table 1).

Activity modification consisted of questions regarding the effectiveness of various activities or lifestyle changes in preventing COVID-19. Policy included questions detailing local or national policy changes enacted in response to COVID-19, including questions about economic support. A full listing of the criteria for each topic category is listed in Table 1.

In line with previous studies, the websites were categorized according to source: World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), non-CDC government, academic, news, and other media [16,17] (Table 1). Specifically, non-CDC government websites consisted of webpages directly maintained by a national governmental entity such as the National Institutes of Health (United States) or the National Health Service (United Kingdom). Academic websites were defined as an organization with a clear academic mission statement. Other media consisted of websites not described by one of the previous categories, including CNET, WebMD, and Wikipedia. A full listing of the criteria for each web source category is listed in Table 1.

Table 1. Question classification by topic and website categorization by source.

Variable	Description
Question classification by topic	
Clinical signs and symptoms	• Properties of SARS-CoV-2, signs and symptoms of COVID-19, prognosis, risk factors, severity, testing • Example: can you have the coronavirus disease without a fever?
Treatment	• Treatment strategies for COVID-19, including antiretrovirals and vaccines; also includes inquiries for unsupported treatments • Example: Can antibiotics treat the coronavirus disease?
Transmission	• Inquiries regarding specific routes of transmission for SARS-CoV-2 • Example: Can coronavirus spread through mosquito bite?
Cleaning methods	• Inquiries regarding specific methods of sanitation to limit spread of COVID-19 • Example: Can ultraviolet light kill the coronavirus disease?
Activity modification	• Questions regarding effectiveness of various activities or lifestyle changes in preventing COVID-19 • Example: Can I walk my dog during quarantine?
Policy	• Questions about local or national policy changes enacted in response to COVID-19; includes questions about economic support • Example: Does everyone get a stimulus check?
Website categorization by source	
World Health Organization (WHO)	• Website maintained by the WHO and hosted on the who.int domain • Example: COVID-19 Situation Report (who.int)
Centers for Disease Control and Prevention (CDC)	• Website maintained by the United States CDC and hosted on the cdc.gov domain • Example: Travelers' health (nc.cdc.gov/travel)
Government	• Websites maintained by national, state, or local governmental organizations other than the WHO or CDC • Example: New York State Governor's Office (governor.ny.gov)
Academic	• Websites maintained by an institution with a clear academic mandate, including universities, academic medical centers, and academic societies • Example: ClevelandClinic.org, MayoClinic.org
News	• Websites published by local, national, or international news media organizations • Example: USAToday.com
Other media	• Websites maintained by organizations not defined in the categories above • Example: Wikipedia.org

Results

In total, 200 questions and their corresponding source of information were extracted; the top 25 questions are listed in Textbox 1. The most common question topic was transmission (n=63, 31.5%), followed by clinical signs and symptoms (n=54, 27.0%) and activity modification (n=31, 15.5%) (Table 2).

Most questions regarding the transmissibility of the coronavirus asked about specific modes of transmission such as spread through food, feces, air conditioning units, delivery packages, and mosquitoes. Interestingly, the clinical signs and symptoms category captured questions about myths associated with the disease, such as whether consuming alcohol stops the coronavirus. In the activity modifications category, there were many questions about staying at home, wearing masks, and managing pre-existing travel plans. The most commonly asked question—"Can antibiotics treat the coronavirus disease?"—was classified as treatment, which, in total, comprised 9% (n=18) of the searched questions.

Textbox 1. Top 25 questions associated with COVID-19.

- Can antibiotics treat the coronavirus disease?
- Is headache a symptom of the coronavirus disease?
- Is there a vaccine for coronavirus?
- Are masks effective against the coronavirus disease?
- Can children pass on the coronavirus disease?
- Can coronavirus spread through mosquito bite?
- Can I stay at home if I have the coronavirus disease?
- Can the coronavirus disease spread faster in an air-conditioned house?
- Can the coronavirus disease spread through delivered packages?
- Can the coronavirus disease spread through feces?
- Can the coronavirus disease spread through food?
- Can you get coronavirus from talking to someone from a distance?
- Can you have the coronavirus disease without a fever?
- Does drinking alcohol kill the coronavirus?
- How long does the coronavirus stay on clothing?
- Is bleach an effective cleaning agent for the coronavirus disease?
- Is the coronavirus disease more severe than the flu?
- Should I accept packages from China?
- Should I cancel my trip due to coronavirus?
- Should I wear a face mask out in public?
- What antiviral drugs are available to treat the coronavirus disease?
- What is a pandemic?
- What is the recovery time for the coronavirus disease?
- Who gets a stimulus check?
- Are rashes a symptom of the coronavirus disease?
- Can hand sanitizer explode in a hot car?

Table 2. Frequencies and percentages associated with questions by topic and websites by source.

Variable	Frequency, n (%)
Questions by topic (n=200)	
Transmission	63 (31.5)
Clinical signs and symptoms	54 (27.0)
Activity modification	31 (15.5)
Policy	22 (11.0)
Treatment	18 (9.0)
Cleaning methods	12 (6.0)
Websites by source (n=200)	
Centers for Disease Control and Prevention	73 (36.5)
Academic	48 (24.0)
World Health Organization	47 (23.5)
News	13 (6.5)
Government	10 (5.0)
Other media	9 (4.5)

With respect to sources of information, the most common websites provided were maintained by the CDC (n=73, 36.5%), academic medical organizations (n=48, 24.0%), and the WHO (n=47, 23.5%) (Table 2). With an additional 5% (n=10) of web information provided by a government source, an overwhelming majority of information (n=178, 89%) came from highly trustworthy web sources. However, the remaining 11% (n=22) of information came from either news or other media. In particular, 4.5% (n=9) of information came from web sources classified as other media, which included potentially erroneous sources of information such as Wikipedia.

Discussion

Principal Findings

In the midst of a highly unprecedented pandemic with significant economic and public health implications, the internet is a crucial source of information for the general public in order to guide their everyday life. As information is changing rapidly and is compounded by fallacies originating from places of authority, we believe that the pandemic highlights the role of physicians in providing patients the most reliable information based on the highest quality of evidence. Thus, the present study effectively characterized the intellectual curiosity of millions of Americans into 6 easily categorizable groups and demonstrated the origin of the general public's sources of information.

Previous studies have examined search and Twitter trends related to the COVID-19 pandemic from regions around the world, including the United States, China, Italy, and Spain [13-15,18-20]. In April, Husnayain et al [14] examined Google search trends in Taiwan, effectively noting that searches for handwashing drastically increased after a perceived face mask shortage in the country. More recently, Rovetta et al [19] examined the Google search trends in Italy, and characterized the most common search terms in the country, including "face mask," "disinfectant," "symptoms of the coronavirus," "health

bulletin," and "vaccines." In the United States, Chen et al [18] examined over 100 million tweets to track social media conversations about the COVID-19 pandemic. However, to our knowledge, no study has examined internet search patterns related to COVID-19 in the United States. This question is of utmost importance for several reasons. First, the United States represents not only the highest COVID-19 burden in the world, but is also a country where recent well-publicized examples of misinformation originated from the head of state [8]. Second, while Twitter effectively captures a significant source of information, it is by no means comprehensive, and the platform appeals to a select audience [21].

Thus, the present study revealed that the most commonly searched criteria about COVID-19 included information about transmission, clinical signs and symptoms, and activity modification. Understanding what matters to our patients should compel us to be well informed on these topics. We believe that as physicians, we should collectively take responsibility to provide reliable information based on the best available evidence [22]. Even if we do not regularly manage patients with COVID-19, at the very least we should be prepared to answer the most common clinical questions asked online such as modes of transmission or the status of a vaccine. Although many answers may be obvious to us, there remain many questions that are active areas of study for which we must remain up to date. Regardless of the specific details of our individual practice, we should always have the willingness to learn and the preparation to answer these commonly asked questions.

Further, the present study revealed that the sizable majority of internet sources provided were from major health organizations or from academic medical institutions. While we find this to be encouraging, we must remember that our patients also consume information from multiple other sources. The social media "echo chamber" phenomenon is an active area of study for sociologists and computer scientists and has been shown to rapidly propagate rumors or misinformation on a mass scale [23-25]. Again, we

believe that physicians should take responsibility for providing the best-quality information in the domains in which we hold influence. On a larger scale, perhaps there is a role for physicians to learn and adapt techniques employed by marketers and politicians to better communicate medical information with the public. However, for most of us, we believe that our role is simply to understand the concerns of our patients regarding COVID-19, to remain informed ourselves, and to be ready to answer their questions.

Limitations

There are several limitations to the present study. First, the COVID-19 pandemic is rapidly changing, and the results of the present study captured web searches as of May 2020. Due to the changes in pandemic characteristics, such as the emergence of new hotspots, it is entirely plausible that the focus of web searches has changed as well. In addition, the present study only captures the web searches of users in New York during May 2020, as the Google COVID-19 database generates search results based on the date of search and user location. Thus, we are unable to analyze trends in either. However, other published studies have examined Google searches in other regions of the world, including China, Taiwan, and Italy [13-15,19]. Thus, the results of the present study should be used in conjunction with those around the world to provide a more comprehensive view of the search patterns of citizens across the world. In addition, the present study makes use of Google's coronavirus homepage, which generates the most commonly asked questions based on the specific user's location and date of search. Due to the limitations of this feature, we are unable to compare trends in location and trends in time, which should be a direction for future studies. Lastly, Google web search was the only search database examined, and the present study fails to capture information from alternative search engines. However, as previously noted, Google is by far the most highly utilized search engine in the United States [9].

Conclusion

People use Google Web Search to identify sources of information about COVID-19. In the United States, the most commonly searched topics related to COVID-19 were transmission, clinical signs and symptoms, and activity modification. Reassuringly, the majority of information in the present study came from highly reputable sources, including the CDC, academic websites, and the WHO.

Conflicts of Interest

None declared.

References

1. Haleem A, Javaid M, Vaishya R. Effects of COVID-19 pandemic in daily life. Curr Med Res Pract 2020 Mar;10(2):78-79 [FREE Full text] [doi: 10.1016/j.cmrp.2020.03.011] [Medline: 32292804]
2. Kache T, Mrowka R. How Simulations May Help Us to Understand the Dynamics of COVID-19 Spread - Visualizing Non-Intuitive Behaviours of a Pandemic (pansim.uni-jena.de). Acta Physiol (Oxf) 2020 Jun 04:e13520 [FREE Full text] [doi: 10.1111/apha.13520] [Medline: 32497348]
3. Vogels EA, Perrin A, Rainie L, Anderson M. 53% of Americans Say the Internet Has Been Essential During the COVID-19 Outbreak. Pew Research Center. 2020 Apr 30. URL: https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-COVID-19-outbreak [accessed 2020-07-09]
4. Koeze E, Popper N. The Virus Changed the Way We Internet. The New York Times. 2020 Apr 7. URL: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html [accessed 2020-07-09]
5. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
6. Fraval A, Ming CY, Holcdorf D, Plunkett V, Tran P. Internet use by orthopaedic outpatients - current trends and practices. Australas Med J 2012;5(12):633-638 [FREE Full text] [doi: 10.4066/AMJ.2012.1530] [Medline: 23382767]
7. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
8. McCarthy T. 'It will disappear': the disinformation Trump spread about the coronavirus – timeline. The Guardian. 2020 Apr 14. URL: https://www.theguardian.com/us-news/2020/apr/14/trump-coronavirus-alerts-disinformation-timeline [accessed 2020-07-09]
9. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
10. Tripathi S, Singh C, Kumar A, Pandey C, Jain N. Bidirectional Transformer Based Multi-Task Learning for Natural Language Understanding. In: Métais E, Meziane F, Vadera S, Sugumaran M, Saraee M, editors. Natural Language Processing and Information Systems. NLDB 2019. Lecture Notes in Computer Science. Cham: Springer; 2019:54-65.
11. Rothwell J. In Mixed Company: Communicating in Small Groups and Teams, 10th ed. New York: Oxford University Press; 2019.

12. Kanthawala S, Vermeesch A, Given B, Huh J. Answers to Health Questions: Internet Search Results Versus Online Health Community Responses. J Med Internet Res 2016 Apr 28;18(4):e95 [FREE Full text] [doi: 10.2196/jmir.5369] [Medline: 27125622]

13. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

14. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]

15. Tao Z, Chu G, McGrath C, Hua F, Leung YY, Yang W, et al. Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]

16. López-Jornet P, Camacho-Alonso F. The quality of internet sites providing information relating to oral cancer. Oral Oncol 2009 Sep;45(9):e95-e98. [doi: 10.1016/j.oraloncology.2009.03.017] [Medline: 19457707]

17. Starman JS, Gettys FK, Capo JA, Fleischli JE, Norton HJ, Karunakar MA. Quality and content of Internet-based information for ten common orthopaedic sports medicine diagnoses. J Bone Joint Surg Am 2010 Jul 07;92(7):1612-1618. [doi: 10.2106/JBJS.I.00821] [Medline: 20595567]

18. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]

19. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]

20. Hernández-García I, Giménez-Júlvez T. Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]

21. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]

22. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]

23. Choi D, Chun S, Oh H, Han J, Kwon TT. Rumor Propagation is Amplified by Echo Chambers in Social Media. Sci Rep 2020 Jan 15;10(1):310 [FREE Full text] [doi: 10.1038/s41598-019-57272-3] [Medline: 31941980]

24. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]

25. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]

Abbreviations

CDC: Centers for Disease Control and Prevention
WHO: World Health Organization
XPath: XML Path Language

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media

Koustuv Saha[1], BTech; John Torous[2], MD; Eric D Caine[3], MD; Munmun De Choudhury[1], PhD

[1]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

[2]Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[3]Department of Psychiatry, University of Rochester, Rochester, NY, United States

Corresponding Author:
Koustuv Saha, BTech
School of Interactive Computing
Georgia Institute of Technology
North Ave NW
Atlanta, GA, 30332
United States
Phone: 1 4046929496
Email: koustuv.saha@gatech.edu

Abstract

Background: The COVID-19 pandemic has caused several disruptions in personal and collective lives worldwide. The uncertainties surrounding the pandemic have also led to multifaceted mental health concerns, which can be exacerbated with precautionary measures such as social distancing and self-quarantining, as well as societal impacts such as economic downturn and job loss. Despite noting this as a "mental health tsunami", the psychological effects of the COVID-19 crisis remain unexplored at scale. Consequently, public health stakeholders are currently limited in identifying ways to provide timely and tailored support during these circumstances.

Objective: Our study aims to provide insights regarding people's psychosocial concerns during the COVID-19 pandemic by leveraging social media data. We aim to study the temporal and linguistic changes in symptomatic mental health and support expressions in the pandemic context.

Methods: We obtained about 60 million Twitter streaming posts originating from the United States from March 24 to May 24, 2020, and compared these with about 40 million posts from a comparable period in 2019 to attribute the effect of COVID-19 on people's social media self-disclosure. Using these data sets, we studied people's self-disclosure on social media in terms of symptomatic mental health concerns and expressions of support. We employed transfer learning classifiers that identified the social media language indicative of mental health outcomes (anxiety, depression, stress, and suicidal ideation) and support (emotional and informational support). We then examined the changes in psychosocial expressions over time and language, comparing the 2020 and 2019 data sets.

Results: We found that all of the examined psychosocial expressions have significantly increased during the COVID-19 crisis—mental health symptomatic expressions have increased by about 14%, and support expressions have increased by about 5%, both thematically related to COVID-19. We also observed a steady decline and eventual plateauing in these expressions during the COVID-19 pandemic, which may have been due to habituation or due to supportive policy measures enacted during this period. Our language analyses highlighted that people express concerns that are specific to and contextually related to the COVID-19 crisis.

Conclusions: We studied the psychosocial effects of the COVID-19 crisis by using social media data from 2020, finding that people's mental health symptomatic and support expressions significantly increased during the COVID-19 period as compared to similar data from 2019. However, this effect gradually lessened over time, suggesting that people adapted to the circumstances and their "new normal." Our linguistic analyses revealed that people expressed mental health concerns regarding personal and professional challenges, health care and precautionary measures, and pandemic-related awareness. This study shows the potential to provide insights to mental health care and stakeholders and policy makers in planning and implementing measures to mitigate mental health risks amid the health crisis.

Introduction

The impacts of global public health emergencies extend beyond medical repercussions; they affect individuals and societies on many levels, causing disruptions [1,2]. In an article written by the American Psychological Association following the Ebola outbreak [3], the epidemic was described as an "epidemic of fear." In the United States, it was labeled by the media as "fearbola" to describe a paranoia that infected the country. Reports of similar feelings of anxiety, stress, and uncertainty have been repeatedly reported during other global outbreaks and pandemics [4]. The ongoing outbreak of SARS-COV-2 has led to a pandemic of illness (COVID-19) that has caused 16 million cases and 700,000 deaths worldwide, reported as of the end of July 2020 [5]. According to recent surveys from the Census Bureau, the Centers for Disease Control and Prevention, and other studies, the COVID-19 crisis has been reported to be associated with rapid rises in psychological distress across many nations [6], with women, the young, the less educated, and some ethnic minority groups reporting greater mental health strain [7]. On the one hand, people are worried about the direct effects of potential infection, including fears of death, lasting disabilities, or exacerbating chronic illnesses. On the other hand, actions to mitigate the spread of COVID-19, including social distancing, quarantines, and business closures with resulting job losses, are a source of life disruptions and emotional distress.

Fear and anxiety about a disease as infectious as COVID-19 can trigger new-onset or exacerbate existing mental illness [8]. Therefore, the practical impact of the crisis is far greater than the actual number of infection cases or fatalities [8,9]. Although expressions of distress may stem from concern and worry relating to the direct impacts of the disease, they may relate as much to disruption of regular routines, sleep and eating patterns, having children out of school and at home full-time, economic hardships and unusual volatility in financial markets, and forced geographical displacement or confinement. Indeed, some people are at risk of developing posttraumatic distress due to exposure to the multifaceted uncertainties or from confronting dying people or lost loved ones. Although disease mitigating efforts such as "social distancing" and "self-quarantining" are recommended [10-12], individuals in medical isolation may experience increased symptoms of anxiety and depression, as well as feelings of fear, abandonment, loneliness, and stigmatization [13,14].

Despite concerns about the myriad of social and behavioral issues associated with the COVID-19 pandemic [15,16], there has been scant research to examine its psychosocial impacts or how to predict and mitigate them. Although it is anticipated that COVID-19 will have broadly ramifying effects [17,18], public health workers and crisis interventionists are limited in their ability to extend services and support in a timely, preemptive fashion. Although surveys are a step forward to support such efforts [7], due to their retrospective recall bias, limited scalability, and inability to provide real-time insights, public health workers are not only unable to prioritize services for the most vulnerable populations but also, more specifically, less equipped to direct prevention efforts toward individuals with greater propensities for adverse psychological impacts.

This paper seeks to address the aforementioned gap by drawing insights into people's expressed mental health concerns by leveraging social media data. The rise in online and social media activity has provided an unprecedented opportunity to enhance the identification and monitoring strategies of various mental and psychosocial disorders [7,19]. Over 80% of US adults use social media daily [20], placing it ahead of texting, email, and instant messaging, and disclose considerably more about themselves online than offline [21,22]. Social media provides a real-time platform where people often candidly self-disclose their concerns, opinions, and struggles during this pandemic [23]. In particular, our study targets the following research aims:

- Aim 1: To quantitatively assess the psychosocial effects of the COVID-19 pandemic using social media data
- Aim 2: To examine how the psychosocial effects of the COVID-19 pandemic have varied over time
- Aim 3: To examine if social media language reflects the major psychosocial concerns during the COVID-19 pandemic

For these research questions, we measured psychosocial effects in terms of symptomatic mental health expressions of anxiety, depression, stress, and suicidal ideation, and expressions seeking emotional and informational support. Our study is founded on a large body of work on studying mental health and psychosocial dynamics with social media data [24-30]. Several studies have leveraged Twitter (which is also the data we used) to study health attributes and public health [30], including symptoms related to diseases [31], disease contagion [32], obesity and physical health [33], mood and depressive disorders [28,34], mental health self-disclosures [27], posttraumatic stress disorder [35], addictive behaviors and substance use [36,37], etc. Because social media data (and Twitter posts in particular) are recorded in the moment, they provide rich information about both the individual as well as the larger world [30]. In particular, we draw on two kinds of prior work: symptomatic mental health expressions and support expressions. Related to the former, Saha et al [25], in their study on the effects of psychiatric medications on Twitter, developed classifiers of mental health symptomatic expressions using social media language, which we replicated in this study. Related to the latter, we draw upon Sharma and De Choudhury's [38], and Saha and Sharma's [39] developed classifiers of social support expressions, specifically emotional and informational support.

Methods

Data

Using Twitter Data

To study people's psychosocial expressions on social media, we obtained Twitter data. Twitter is one of the most popular social media platforms, and its public-facing, microblogging–based design enables candid self-disclosure and self-expressions for individuals [27].

Twitter Streaming Application Programming Interface

We collected data in our study using the Twitter Streaming application programming interface (API). The Twitter Streaming API is an official data collection API that Twitter shares with researchers providing free access to a 1% sample of its data on parameters set by researchers. That is, for a given set of parameters, Twitter queries the volume of available data at a particular moment [40]. If the volume of the query exceeds 1% of all Twitter posts at that moment, then the response is sampled to be less than 1%. However, the Twitter Streaming API is like a black-box with a lack of transparency in the sampling methodology [41,42], yet this is one of the few forms of unfettered and large-scale social media data access to researchers outside social media companies [43] and has been used in prior research, including in health-related studies [35,44,45]. The Discussion section revisits the limitations of our study due to the challenges of the Twitter Streaming API.

For the purposes of our study, we used two kinds of parameters: (1) language of a Twitter post as "english" and (2) geolocation bounds set to be within the geographic coordinates of the United States. Therefore, our following analyses concern Twitter data that at least fulfill both these criteria. We note that the location filter additionally prevents any retweets in the data set, as retweets are not geolocation labeled by design on Twitter [40], allowing us to study only originally created Twitter posts.

Treatment Data

In particular, we focused our study on the US population and leveraged the Twitter Streaming API. Using geo-bounded coordinates, we collected 1% of real-time Twitter data originating from the United States. We collected 59,096,694 Twitter posts between March 24, 2020, and May 24, 2020. Because this data set comes from the same period when the COVID-19 outbreak occurred, we labeled this data set as the *treatment* data set. We note that this period saw an exponential growth in reported COVID-19 infection cases (about 50,000-1 million) and fatalities (about 1000-56,000) in the United States [46]. During these 2 months, federal and state policies and laws were enacted to control or mitigate the spread of the outbreak, including school and work closures; stay-at-home orders; and the Coronavirus Aid, Relief, and Economic Security Act [47].

Control Data

To understand the social media expressions particularly attributed to the COVID-19 crisis, we obtained a control data set that originated from the same geographical location (the United States) and a similar time period but from the previous year (2019). Prior work [47,48] motivated this approach of obtaining control data that acts as a baseline and likely minimizes confounding effects due to geo-temporal seasonality in lifestyle, activities, experiences, and unrelated events that may have some psychosocial bearing. We obtained a similarly sized data set of 40,875,185 Twitter posts shared between March 24, 2019, and May 24, 2019.

Both the treatment and control data sets were collected in real time, and therefore, they were the entire 1% sample of Twitter posts returned in real time; we did not conduct any additional sampling on this data. We note that the size of the control data was smaller than that of the treatment [42] despite each consisting of the same 2-month duration. This could be because the volume of posts [40,42] on Twitter increased significantly in 2020 [49], leading to an increase in the 1% sample as well. However, we cannot make any such conclusion, especially because of the lack of transparency in how Twitter conducts the 1% sampling [42].

Psychosocial Effects of COVID-19

Study Design

Toward our first research aim of understanding the psychosocial impacts of the COVID-19 outbreak, we conducted two types of analysis on our Twitter data set, which we describe in the following sections. Our study builds upon the vast, rapidly growing literature studying mental health concerns and psychosocial expressions within social media data [19,21,24-28,34,48,50-52]. We adopted a quasi-experimental study design, which minimizes for geo-temporal confounds by using Twitter data sets from similar regions and similar times of the year in a treatment (2020) and a control (2019) year.

Symptomatic Mental Health Expressions

Drawing on the work previously referenced, we hypothesize that people's self-disclosure expressions on social media can reveal symptomatic mental health expressions attributed to the COVID-19 crisis. We examined symptomatic expressions of anxiety, depression, stress, and suicidal ideation. These are not only some of the most critical mental health concerns but also have been attributed to be consequences of the pandemic outbreak [15,53,54].

To identify mental health symptomatic expressions in social media language, Saha et al [25] built machine learning classifiers using transfer learning methodologies—the main idea here is to infer mental health attributes in an unlabeled data by transferring a classifier trained on a different labeled data set. These classifiers are *n*-gram–based (*n*=1,2,3) binary support vector machine (SVM) models, where the positive class of the training data sets stems from appropriate Reddit communities (*r/depression* for depression, *r/anxiety* for anxiety, *r/stress* for stress, and *r/SuicideWatch* for suicidal ideation), and the negative class of training data sets comes from nonmental health–related content on Reddit—a collated sample of 20 million posts gathered from 20 subreddits from the landing page of Reddit during the same period as the mental health subreddit posts, such as *r/AskReddit*, *r/aww*, and *r/movies*. These classifiers perform at a high accuracy of approximately 0.90 on average on held-out test data [25].

Clinical Validity

Saha et al's [25] classifiers used here have also been shown to transfer well on Twitter with an 87% agreement between machine-predicted labels and expert appraisal [48], where experts annotated posts in the classification test data using Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [55] criteria of mental health symptoms. Bagroy et al [56] reported additional validation of such derived insights with feedback from clinical experts [55]. In this study, the outcomes of the mental health expression classifiers were compared with those given by human coders on the same (random) sample of social media posts; the latter coded the posts based on a codebook developed using prior qualitative and quantitative studies of mental health disclosures on social media and literature in psychology on markers of mental health expressions.

Coders not only agreed with the outcomes of the classifiers (κ=0.83) but also noted that the classifiers could identify explicit expressions of firsthand experience of psychological distress or mental health concerns ("i get overwhelmingly depressed") as well as expressions of support, help, or advice around difficult life challenges and experiences ("are there any resources I can use to talk to someone about depression?"). Further details about these classifiers, including their detailed performance, predictive features demonstrating model interpretability, and efficacy of transfer to Twitter data, can be found in [25,48,56]. We used these classifiers to machine label both our *treatment* and *control* data sets. Textbox 1 shows example Twitter posts in our data set that exhibit mental health symptomatic expressions (because many of these labels were comorbid, we show example posts that exhibit one or more of these mental health symptomatic expressions).

Textbox 1. Example paraphrased posts in the treatment data that exhibited high symptomatic mental health expressions.

- I am so sick and tired of the #coronavirus (anxiety, stress)

- The kind of person I am, I don't deserve to meet these people (depression)

- 2020 is the saddest year. There is a lack of money, necessities needed for daily life are gone from stores! I am at work as a healthcare professional hurting for my patients because they can't see their family. (anxiety, stress)

- I am too overwhelmed by school and having a crippling anxiety to keep up with everything online I'm seriously NOT OK! I CAN'T TAKE THIS! (anxiety, suicidal ideation)

- During the online lecture, the prof. asked each of us how we were doing and feeling, I said, I am too anxious to know what's next, and i keep thinking about what's gonna happen, and she was like okay chill! (depression, anxiety, stress)

Support Expressions

Social support is considered an essential component in helping people cope with psychological distress [57]. Research reports that supportive interactions can even have a "buffering effect" [57,58]; that is, they can be protective against the negative consequences of mental health. With the wide adoption of web and social media technologies, support seeking (and providing) is increasingly happening online and has been shown to be efficacious [21,59]. In fact, a meta-analysis indicated that online support is effective in decreasing depression and increasing self-efficacy and quality of life [60]. In the context of suicide, certain types of social support in Reddit communities may reduce the chances of future suicidal ideation among those seeking mental health help [61]. Oh et al [62] further showed that surveyed Facebook users demonstrated a positive relationship between having health concerns and seeking health-related social support. Indeed, during global crises such as COVID-19, when many of the physical sites for health care (including mental health) have been closed or have restricted access, it is likely that online support has proliferated [63]. Fear of potential infection may further have alienated individuals in need to pursue formal treatment, therapy, and support, perhaps channelizing their support-seeking efforts online and on social media.

According to the "Social Support Behavioral Code" [64], two forms of support that have received theoretical and empirical attention are emotional and informational support. Emotional support (ES) corresponds to empathy, encouragement, and kindness, while informational support (IS) corresponds to

information, guidance, and suggestions [38,65]. These two forms of support have been found to be most prevalent and effective in several studies of online support and social media [38,62,66,67]. Social media enables individuals to self-disclose and express their emotional and informational needs [67]. Andalibi et al [66] found that these two kinds of support can co-occur with other forms of support, such as posts seeking ES often seeking esteem and network support, and Attai et al [68] noted that Twitter is effective in seeking and providing health-related informational needs, contextually related with our problem of interest.

To identify support expressions on social media, we used an expert-appraised data set and classifier built in prior studies [38,39]. These are binary SVM classifiers identifying the degree (high and low) of ES and IS in social media posts. When the predictions of these classifiers were cross-validated with expert annotations from Sharma and De Choudhury's [38] data, the classifiers were found to have *k*-fold cross-validation accuracies of 0.71 and 0.77 in ES and IS classifications, respectively [39]. Similar to the symptomatic expressions classifiers, the classifiers of support expressions are transferred from Reddit and typically performed well in our data set due to the high linguistic equivalence between Reddit and Twitter data sets [34]. We further manually inspected a random set of 125 Twitter posts in our data set using the methods outlined in prior studies [25,56] to rate each Twitter post with binary high or low ES and IS. We found that the manual ratings and classifier ratings showed a high agreement of 88% and 93%, respectively, indicating statistically significant transfer classification on Twitter. We used these classifiers to label the presence of ES and IS in our

treatment and *control* data sets. Textbox 2 shows a few example paraphrased posts of support expressions in our *treatment* data set.

Textbox 2. Example paraphrased posts in the treatment data on support expressions.

- To our residents, the town is here for you but we need your help if we are going to keep our hospitals from being overwhelmed. (emotional and informational support)

- I hope you are safe and healthy! Keep the faith God never fails us & always has perfect plan. (emotional support)

- my wife is laid off. She has been trying for days to get ahold of unemployment. Please help. We need income. Running low on basic things. (emotional and informational support)

- Dear God, we are going through some struggles these days. Could you please send us some sunshine? Thanks so much. (emotional support)

- According to my mom, kindness is needed more than ever now. So, send love to you! How are you being kind to others today? (emotional support)

Examining Psychosocial Expressions Over Time and Language

Approach Overview

Next, we describe the methods to examine how the COVID-19 pandemic may have caused changes in psychological expressions by comparing our *treatment* (outbreak year) and *control* (no outbreak year) data sets. For both our data sets, we aggregated the number of posts that expressed symptomatic and support expressions by day and by type. We compared the pervasiveness of each kind of measure in the data sets along with conducting statistical significance in their differences using two-sample *t* tests and effect sizes (Cohen *d*).

Temporal Variation

For our second research aim, we compared the daily variation of measures between the *treatment* and *control* data sets, we transformed our data into standardized *z* scores. Our data sets relied on the Twitter Streaming API and were subject to daily inconsistencies of available data each day [41]. Transformed *z* scores are not sensitive to such absolute values and inconsistencies, and essentially quantify the number of SDs by which the value of the raw score is above or below the mean. Similar standardization techniques have been adopted in prior social media time series studies [48,69]. *Z* scores were calculated as $(x - \mu) / \sigma$, where *x* is the raw value, μ is the mean, and σ is the SD of the population. Here, to obtain population μ and σ, in addition to our *treatment* and *control* data, we also included a year-long Twitter data of over 240 million Twitter posts (September 2018 to August 2019). For each of the measures in symptomatic and support expressions, we first calculated μ and σ on the per-day occurrence of that particular measure in the data set of over 300 million Twitter posts (combining 240 million posts between September 2018 and August 2019, and 60 million posts in the treatment data between March and May 2020). For each measure, we then calculated the *z* score per day and interpreted the positive *z* scores as values above the mean and negative *z* scores as those below the mean.

Linguistic Differences

For our third research aim, we examined COVID-19–related linguistic differences in the psychosocial expressions on social media, we employed an unsupervised language modeling technique, the Sparse Additive Generative Model (SAGE) [70].

Given any two data sets, SAGE selects salient keywords by comparing the parameters of two logistically parameterized-multinomial models using a self-tuned regularization parameter to control the trade-off between frequent and rare keywords. We conducted SAGE to identify distinguishing *n*-grams (*n*=1,2,3) between the *treatment* and *control* data sets, where each *n*-gram was returned with a SAGE score. The magnitude of an *n*-gram's SAGE score signals the degree of its "uniqueness" or saliency, and in our case, a positive SAGE score (above 0) indicated that the *n*-gram was more salient in the *treatment* data, whereas a negative SAGE score (below 0) denoted greater saliency in the *control* data.

SAGE allowed us to obtain how the expressions differ during the COVID-19 outbreak as compared to the *control* period. We conducted two SAGE analyses, one each for symptomatic and support expressions. For the symptomatic expressions, we first obtained posts that were indicative of either anxiety, depression, stress, or suicidal ideation in the *treatment* and *control* data sets, and obtained SAGE for both. We used a similar method for support expressions by obtaining posts that were indicative of either emotional or informational support.

Finally, we cross-examined the salient keywords across symptomatic and support expressions to study how concerns were prevalent in either or both of the expression types. We measured log-likelihood ratios (LLRs) along with add-one smoothing, where LLRs close to 0 indicated comparable frequencies, LLRs<1 indicated the greater frequency in symptomatic expressions, and LLRs>1 indicated the greater frequency in support expressions. Together, these linguistic analyses enabled us to obtain psychological concerns and understand how COVID-19 has psychosocially affected individuals, and to contextualize these concerns in the literature on consequences of global crises.

Results

Summary of Results

We summarize our first set of results in Table 1. For all our measures, we found statistical significance (as per *t* tests) as well as significant effect sizes (Cohen *d*>0.4 for all measures [71]) in social media expressions in the *treatment* data as compared to that in the *control* data. Assuming that most other confounders were minimized due to the geo-temporal similarity

of the data sets, our findings indicated that the COVID-19 outbreak led to an increase in people's symptomatic and support expressions of mental health. We elaborate on the results in the following sections.

Table 1. Comparing social media expressions in the treatment (2020) and control (2019) data sets.

Expression	Treatment (2020), mean (SD)	Control (2019), mean (SD)	Δ (%)	Cohen d	t test (df)	P value
Symptomatic mental health expressions						
Anxiety	1.65 (0.20)	1.35 (0.08)	21.32	1.96	12.60 (151)	<.001
Depression	9.00 (0.60)	8.17 (0.35)	10.18	1.71	10.72 (151)	<.001
Stress	19.31 (0.77)	18.61 (0.43)	3.76	0.81	3.65 (151)	.009
Suicidal ideation	3.14 (0.31)	2.62 (0.13)	19.73	2.14	13.54 (151)	<.001
Support expressions						
Emotional support	8.56 (0.84)	8.17 (0.50)	4.77	0.46	2.87 (151)	.004
Informational support	1.75 (0.18)	1.67 (0.08)	4.78	0.56	3.58 (151)	<.001

Temporal Variation

Figure 1 shows the changes in symptomatic mental health expressions for the same period in the *treatment* (2020) and *control* (2019) years. We found that the *treatment* and *control* data sets showed significant differences in people's symptomatic expressions (Table 1), among which anxiety showed the most significant increase (21.32%), followed by suicidal ideation (19.73%), depression (10.18%), and stress (3.76%). Figure 2 shows the evolution of support expressions change in the *treatment* and *control* data sets. The differences were significant (Table 1), and we found that ES increased by 4.77% and IS increased by 4.78%.

Figure 1. Comparison of symptomatic mental health expressions on social media posts in the same period (March 24 to May 24) in 2019 and 2020 (COVID-19 outbreak year). Idn.: ideation.
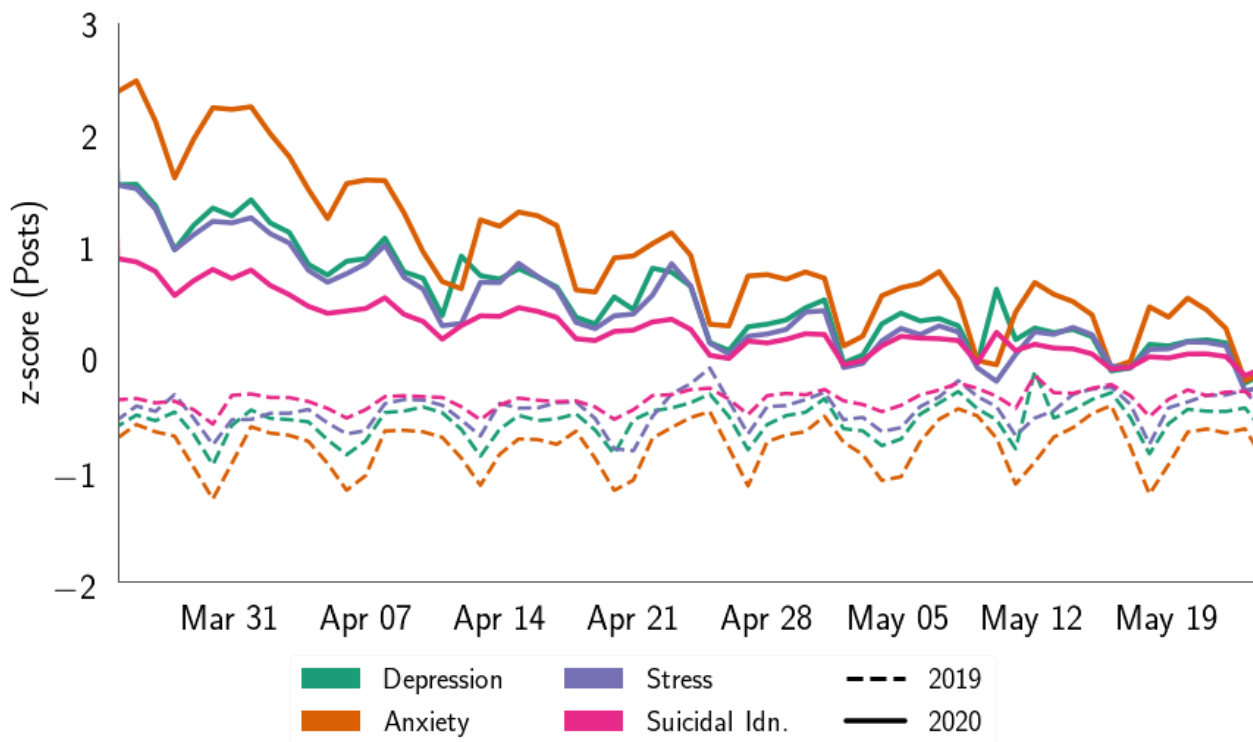
Figure 2. Comparison of support expressions on social media posts in the same period (March 24 to May 24) in 2019 and 2020 (COVID-19 outbreak year).
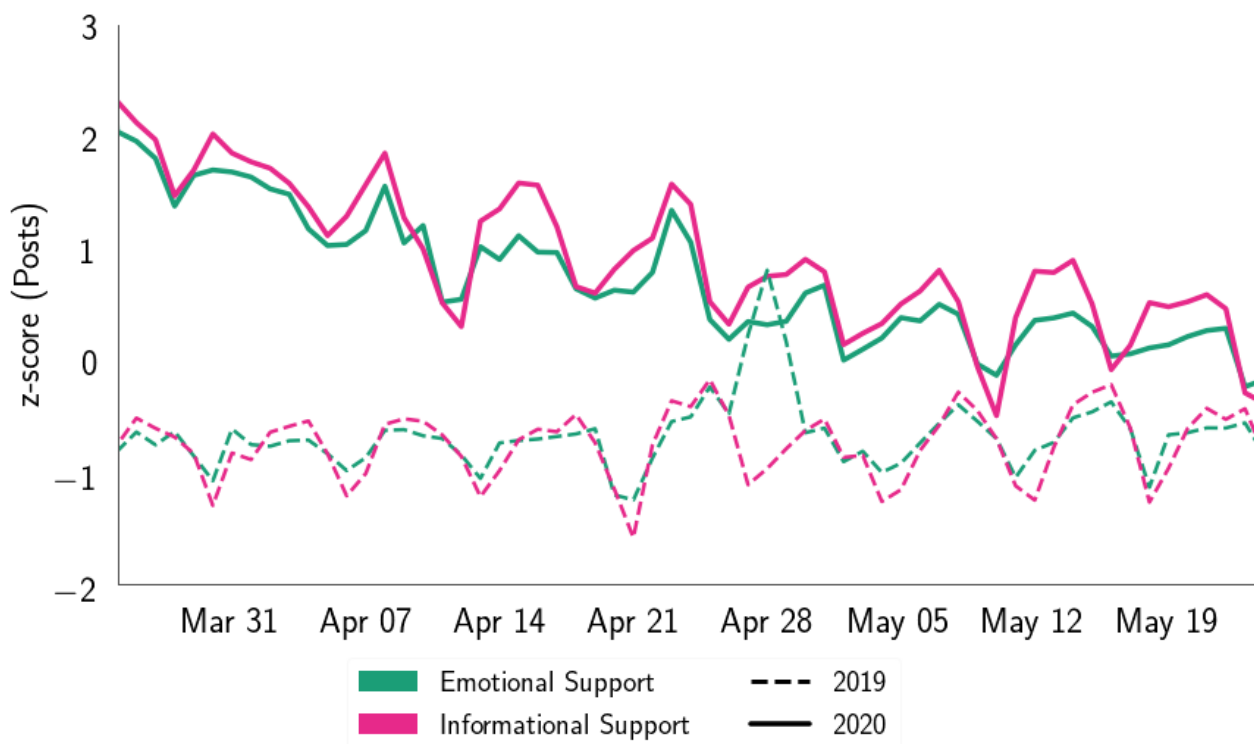
In both the plots of Figures 1 and 2, we found a general trend of negative slope (average slope=–0.03) within the *treatment* year, which was closer to zero slope (average slope=$3.19*10^{-4}$) in the *control* data set. This may suggest that within the *treatment* year, people's mental health expressions gradually leveled out over time, despite the growing rate of COVID-19 active cases. The plots indicated that psychological expressions almost converge at the tails. This could likely be due to people's habituation with the situation and surroundings with the passage of time [70,72], as has been observed for other crisis events [48,73]; however, this needs to be explored further. Within the *control* data set, we observed a sudden peak on April 28, 2019, which could be attributed to a shooting incident at a synagogue in San Diego [74]. The observations reflected that the COVID-19 pandemic has increased people's mental health expressions on social media, aligning with other contemporary literature and media reports [8,53].

Linguistic Expressions

Symptomatic Mental Health Expressions

Table 2 summarizes the language differences as per SAGE for posts expressing high mental health expressions in the *treatment* and *control* periods—keywords with positive SAGE saliently occurred in the *treatment* data, whereas those with negative SAGE saliently occurred in the *control* data. A majority of the keywords that occurred in the *treatment* period were contextually related to the COVID-19 pandemic, such as *covid19*, *coronavirus*, *social distancing*, and *stayathome isolation*. These keywords were used in posts expressing mental health concerns either explicitly (eg, "Social distancing is both sad and anxiety-inducing at the same moment") or implicitly (eg, "In order to get my family treated, I will do more than beg, and I will donate 25K for research to develop COVID19 vaccine"). We also found that the *treatment* period used keywords referring to key personnel such as *dr fauci* (referring to Anthony Fauci, one of the leads in the incumbent White House Coronavirus Task Force in the United States and Director of the National Institute of Allergy and Infectious Diseases since 1984 [75]) and political figures like *Nury Martinez* and *Donald Trump*. Further, we found keywords such as *essential workers*, *doctor jobs*, and *risking lives*, which describe high-risk worker situations (eg, "I am not complaining about going to work, rather, I am concerned about risking my health for work"), and certain treatment suggestions that evolved during this period [76], such as *garlic*, *malaria*, and *hydroxychloroquine* (eg, "I do worry tho! He is focused on job numbers, approval? ratings and repeating mistruths. His spouting of 2 drugs, one for malaria & the other a Z-pack. A Senior couple tried these untested drugs; wife is in ICU & husband also hospitalized! This is irresponsible & dangerous!").

Table 2. Top salient n-grams (n=1,2,3) for symptomatic mental health concerns in the treatment and control data sets (SAGE analysis [70]).

Salient in treatment (2020)		Salient in control (2019)	
Keyword	SAGE[a,b]	Keyword	SAGE[c]
covid19	11.17	hospitality	−2.81
lord marvelous	10.87	trainee	−2.78
coronavirus	10.58	crimes	−2.74
social distancing	9.92	delay	−2.55
nury martinez	9.66	traffic	−2.55
working councilwoman	9.66	accident	−2.39
bored daily	8.69	finance accounting	−2.26
stayathome isolation	8.69	half finance	−2.22
quarantinelife	8.62	auburn	−2.21
quarantine got	7.87	half technology	−2.19
securityguard	7.63	pete	−2.19
essential workers	7.62	parttime	−2.18
dr fauci	7.56	robert half	−2.12
went tired	7.48	tickets	−2.08
coronaviruspandemic	7.44	marvel	−2.07
flattenthecurve	7.44	tournament	−1.92
doctorjobs	7.32	muslim	−1.90
garlic	7.29	florida	−1.90
hydroxychloroquine	7.28	boston	−1.88
n95 masks	7.26	cashier	−1.87
masks gloves	7.26	playoffs	−1.86
practice social distancing	7.21	sales representative	−1.85
physicianjobs doctorjobs	7.13	springfield	−1.84
quarantine life	6.98	border	−1.84
plz help small	6.96	barista	−1.77
small donation	6.96	israel	−1.77
stay home orders	6.92	nc click	−1.76
self quarantine	6.88	playoff	−1.75
positive covid19	6.79	bracket	−1.75
risking lives	6.79	terrorist	−1.65

[a]SAGE: Sparse Additive Generative Model.

[b]Positive SAGE scores indicated greater saliency in the treatment (2020) data.

[c]Negative SAGE scores indicated greater saliency in the control (2019) data.

Support Expressions

Table 3 lists the top keywords as per SAGE for support posts in the *treatment* and *control* periods. Keywords with positive SAGE saliently occurred in the *treatment* data, whereas those with negative SAGE saliently occurred in the *control* data. We found keywords that explicitly relate to COVID-19 occurred in the *treatment* period. We also found that the *treatment* period consisted of posts that seek support related to job and pay, such as *losing jobs*, *need pay*, and *furloughed* (eg, "Many in our

community have lost their jobs, are underinsured and are struggling to make ends meet. Providing pantries, hot meals, hotspots and distance learning opportunities is now more critical than ever, please donate"). Our data also revealed the prevalence of contextually related keywords such as *masks*, *ppe*, *hoarding*, *stockpile*, and *sanitizer* that are medically recommended prevention and containment measures of COVID-19 infection (eg, "Please contact me if you have any N95 mask or know to obtain some. My sister and a few friends work in the OR and

they do not have the supplies to stay safe, they have patients who have #COVID19. TY! #HealthcareHeroes").

Table 3. Top salient n-grams (n=1,2,3) for support expressions in treatment and control data sets (SAGE analysis [70]).

Salient in treatment (2020)		Salient in control (2019)	
Keyword	SAGE[a,b]	Keyword	SAGE[c]
lord	7.93	hospitality	–2.86
fauci	6.70	duke	–2.51
ventilators	6.59	shift supervisor	–2.24
quarantine	6.47	tampa	–2.21
securityofficer	6.11	advisor	–1.95
n95	5.53	customerservice	–1.92
hope staying safe	5.36	investigation	–1.89
ppe	5.25	manager retail	–1.87
wearing masks	5.20	traffic	–1.87
uncertain times	5.16	muslim	–1.86
healthcare workers	5.01	store manager	–1.85
furloughed	5.00	tickets	–1.85
asymptomatic	4.95	playoffs	–1.83
people quarantine	4.90	cubs	–1.82
fighting stigma	4.82	border	–1.81
staysafe	4.67	cashier	–1.79
food bills	4.66	springfield	–1.79
disinfectant	4.64	delay	–1.76
hand sanitizer	3.08	barista store	–1.76
clorox	3.03	boston	–1.76
medical supplies	2.97	counter	–1.75
trying times	2.89	barista	–1.74
risking lives	2.87	columbia	–1.73
stockpile	2.86	meeting retail	–1.73
father passed	2.36	informational meeting	–1.73
hoarding	2.31	stlouis	–1.72
mask	2.31	marvel	–1.70
medical professionals	2.27	marketing	–1.68
losing jobs	2.27	server	–1.67
toilet paper	2.05	accident	–1.64

[a]SAGE: Sparse Additive Generative Model.

[b]Positive SAGE scores indicated greater saliency in the treatment (2020) data.

[c]Negative SAGE scores indicated greater saliency in the control (2019) data.

Linguistic Comparability

Finally, Table 4 shows the results of the lexical comparability analysis, where LLRs demarcate the top keywords used for symptomatic mental health expressions and support expressions within the *treatment* data set. We found that keywords, such as safety precautions (*wear masks*), health care and treatment (*health care workers, hospitalized, beds,* and *icu*), and life and death (*passed away, kill people, human lives,* and *deaths*), comparably overlapped in both kinds of psychological expressions (LLRs~0). These keywords were also used to raise awareness and express solidarity with health care and high-risk workers (eg, "Taking all safety precautions and adhering to the guidelines established by our health care professionals will keep us safe"). Our lexico-psychological analyses revealed that more clinically relevant keywords and symptoms occurred frequently in symptomatic expressions (LLRs>0; eg, *sleep schedule* and

tested positive), whereas socially relevant and stressful circumstances were more prevalent in support expressions (LLRs<0; eg, *im single parent, starve,* and *lost jobs*).

Table 4. Distribution of social media keywords across high symptomatic mental health and support expressions within the treatment period using LLRs.

LLRs[a]>0[b]		LLRs~0[c]		LLRs<0[d]	
Keyword	LLR	Keyword	LLR	Keyword	LLR
sleep schedule	0.75	infected	–0.01	im single parent	–1
lonely	0.64	wear masks	–0.01	starve	–1
anxiety	0.62	need help	–0.01	meditate	–1
isolation	0.56	kill people	–0.01	sorry loss	–0.73
stay safe	0.56	need pay	0	care people	–0.7
bored	0.56	health care workers	0	hard times	–0.45
tested positive	0.52	passed away	0	people sick	–0.4
quarantine life	0.52	seriousness	0	helping people	–0.4
homeschooling	0.51	human lives	0	sorry hear	–0.39
tired	0.5	deaths	0	urged	–0.33
doctor	0.48	domestic violence	0	new yorkers	–0.29
fighting stigma	0.46	comforting	0	lost jobs	–0.21
depression	0.45	hospitalized	0	hope family	–0.21
stuck inside	0.42	beds	0	selfish	–0.21
sane	0.41	icu	0.01	desperate	–0.21

[a]LLR: log likelihood ratio.

[b]Keywords with LLRs>0 distinctly occurred in high symptomatic expressions.

[c]Keywords with LLRs~0 occurred comparably in both symptomatic and support expressions.

[d]Keywords with LLRs<0 distinctly occurred in support expressions.

Discussion

Principal Results

Our study suggests that social media posts during the COVID-19 pandemic contain a significantly higher frequency of symptomatic mental health and support expressions than a comparable data set from the same period in the previous year. The effect sizes and statistical differences observed in our analyses provide evidence that COVID-19 may have led to mental health concerns compared to other normative times. We also found that they topically relate to the ongoing crisis situation and include concerns such as treatment, precautionary measures, loss of jobs, school closings, stockpiling of basic livelihood necessities, feeling lonely, boredom, and tired of the restrictions and constraints put on by the ongoing pandemic. Our findings suggest that although the COVID-19 pandemic has amplified mental health risks and concerns, it may have heightened a sense of belonging and solidarity among individuals—bringing them together, raising collective awareness, and encouraging them to provide support to one another. We also found expressions of solidarity with health care and high-risk workers, suggesting that people have been considerate about these workers and have expressed desire to set up opportunities for donating to those who have lost jobs during the crisis; this also aligns with recent media reports and

World Health Organization guidelines of tackling the pandemic [77,78]. Media reports have also indicated how benevolent neighbors have been, tending to their older adult neighbors by delivering their groceries and other basic necessities [79].

However, mental health experts say that, although the crisis is amplifying risk factors for suicide, the COVID-19 outbreak's effect on individuals' mental and emotional well-being is complex [80]. Suicide is multifaceted, and although economic loss is a risk factor, so are depression, isolation, and fear of the future. At the same time, the crisis is possibly creating a sense of belonging for individuals at risk for suicide, as stress and anxiety are normalized, and people come together to better support one another during a crisis [81,82]. Our data showed a significant impact of COVID-19 on suicidal ideation, which calls for enhanced importance of population-scale mental health care, such as using approaches like universal screening (ie, Zero Suicide Model) [83]. As Florida [84] noted in a recent article, "The long-term toll on mental health of social isolation, remote work, and economic insecurity could have impacts akin to post-traumatic stress disorder; yet, the new focus on mental health may reduce stigma and increase the availability of support services." Indeed, the world beyond the crisis may be one in which mental health is more honestly recognized and supported.

Interestingly, we noted that our findings indicate a gradual leveling out of these expressions, both symptomatic and

supportive, which may reflect a developing *new normal*. In February 2020, it seemed unthinkable that the white-collar workforce of many countries would soon be working solely from home; it seemed unthinkable that air travel would plummet by 96% and that all major sporting events would be called off. Indeed, epidemiologists surmise that many if not most of the changes surrounding the rhythms of our daily life are likely to fade over time, just as they did after the 1918 influenza epidemic [84]. In other words, the pandemic could make us revisit and possibly reform many of our lifestyle choices and civic roles, and the persistent discussion of the *new normal* may help bring order to our current turbulence. Others have argued that perhaps the crisis is a prelude to a *new paradigm*, as recently noted by the World Economic Forum [84,85]: "Feeling unsettled, destabilized and alone can help us empathize with individuals who have faced systematic exclusions long-ignored by society even before the rise of COVID-19 – thus stimulating urgent action to improve their condition." We should, therefore, "revel in the discomfort of the current moment to generate a 'new paradigm,' not a 'new normal.'" The leveling out trend in our data gives empirical ground to these conjectures.

Nevertheless, if robust antiviral treatments are developed and rolled out relatively quickly or if a vaccine becomes available soon enough, presumably, the changes will be short-lived, and the new normal may be temporary. However, if the pandemic comes back in larger waves over the next few seasons, as was the case with historical epidemics, the economic, political, and social crises that have arisen as a consequence will lead to deeper ramifications in turn, leading to longer lasting or permanent changes. Future research will need to explore the persistence of the new normal and the emergence of a possible new paradigm as the pandemic evolves, and therein the mental health impacts further along in the crisis. A study like ours on the ongoing pandemic is a step toward leveraging large-scale online data to understand people's response to the crisis and thinking about means to address the major concerns. Our study bears implications in digital technology driven mental health interventions to provide tailored support to people's concerns during the crisis; a recent work by Rudd and Beidas [86] pointed out four point guidelines to build innovative and expansive solutions toward improving public mental health. The variety of concerns and help-seeking factors reflected in our study can also help several stakeholders, ranging across mental health facilitators and policy makers, toward early preparedness and interventions for mental health support. Similarly, our methods can offer the potential to build public health surveillance technologies that surface early warning signs of the effects of the various events related to the pandemic and other crises. The potential of social media to assist in the response to the pandemic is clear but also dependent on the accuracy of underlying methods. The reach of social media allows for broad access that transcends national borders or cultural differences. Using this access to meet the increasing need for help seeking, online and social media data is in a prime position to offer people personalized guidance toward accurate information, health care resources, and even basic lifestyle interventions. Underlying this potential is, thus, the ability of social media data to classify the state and needs of each individual and use that information to tailor a customized response. Precedents for

such a system are abound as seen in several prior studies [27-31,37,87-89].

Comparison With Prior Work

COVID-19 is not the first pandemic—catastrophic pandemics have been occurring at regular intervals throughout human history, with the 1918 influenza epidemic being the last one before the current pandemic [84,85,90]. The backdrop of the 1918 pandemic was that it happened just before the advent of modern psychiatry as a science and a clinical specialty—a time when psychoanalysis was gaining recognition as an established treatment within the medical community [91,92]. Consequently, psychiatry has had little opportunity to consider such historically important phenomena through its clinical, scientific lens until now. Although outbreaks of the Zika and Ebola viruses, Middle East respiratory syndrome, and severe acute respiratory syndrome managed to draw global attention, stirring up anxiety and uncertainty in societies, scholars have noted that participation of mental health experts in pandemic preparedness has remained negligible [93]. Consequently, our ability to understand mental health responses as well as the mental health burden in pandemic outbreaks have been limited [94]. For instance, a routinely practiced method of infection control, quarantine and social distancing, have received surprisingly little attention in psychiatric literature so far. Baumeister and Leary [95] contended that humans need frequent contacts, and crisis events further stimulate a need for affiliation and intimacy. Therefore, prolonged isolation and separation from families and their community can have profound effects on individuals even if they are not directly affected by the disease [4]. In the current pandemic, the additional layer of extensive social media use and exposure to often sensationalized online news while in physical isolation may add new complexities to implementing emotional epidemiology in managing concerns, fears, and misconceptions [96,97], as these tools have been argued to bear negative effects on psychological well-being [97,98].

By adopting social media as a lens to unpack these previously less understood dimensions of a pandemic's mental health effects, our study is one step toward closing some of the previously noted gaps. The published literature posits that the distress and anxiety among individuals in this COVID-19 pandemic may increase the incidence of mental disorders [53,54,99]; data thus far from the United States point to a population increase in psychological distress of 10% compared to 2018 data [8], a trend that is in line with our results. These rates may be higher in those regions heavily exposed to COVID-19 or among individuals working during the pandemic, with a recent review reporting over 20% prevalence of anxiety, which is also consistent with our findings [8].

Prior work found that mental health discourse on Twitter ranges across stigmatizing, inspirational, resource, medical, and social dimensions of expressions [100], and our study revealed similar topical diversity in our data set. Further, we detected through social media many of the stresses associated with the pandemic (eg, prolonged isolation, exposure to pandemic-related death, loss of income and career, increased workload, and lack of pertinent and accurate information). These results align with epidemiological findings that COVID-19 has led to elevated

mental health symptoms for individuals. Nelson et al [101] surveyed 2000 individuals from the United States, Canada, and Europe, and found elevated symptoms of anxiety and depression compared to historical norms and observed factors similar to the concerns we detected regarding symptomatic expressions and those related to seeking support. They also reinforced the summary data released by the Crisis Text Line (a major crisis helpline in the United States) listing major concerns of crisis support sought during this period [102], with 80% of conversations mentioning *"virus,"* 34% mentioning *"anxiety,"* 34% feeling solidarity with friends and family, etc. Along similar lines, there have been numerous reports about the increasing number of mental health crisis helpline calls during this period [103,104], providing further support and external validation that our social media findings reflect many of the same elements of distress expressed offline during this crisis.

Next, our temporal analyses pointed to a steady decline in people's expressed psychosocial concerns during the 2-month study period (Figures 1 and 2), which conforms with similar findings in Google search queries as stay-at-home orders and other COVID-19–related policy changes were implemented in the United States [105]. We note contemporary social computing research studying various aspects of the social media discourse related to COVID-19 [63,106-108]. By providing complementary evidence to observations by Mackey et al [106] and Stokes et al [107] on expressed (mental health) concerns during the crisis, our study further underscores their findings using a comparable (control) data set, reinforcing and providing empirical credibility to the impression that the COVID-19 pandemic has indeed caused or contributed directly to the mental health concerns that we describe.

Limitations and Future Work

We note some limitations in our study, many of which present excellent directions for future research. We recognize the lack of transparency related to the Twitter Streaming API. Recent research has also questioned the credibility of the "1% Twitter stream" aspect, noting that actual sampling data was smaller than what it ideally should have been [41]. Given these data limitations, we decided against conducting several descriptive and fine-grained analyses (such as comparing regions), and refrained from making claims based on comparing absolute numbers of those impacted by various mental health concerns. For example, we cannot define based on our data whether there were increased or decreased Twitter postings during our COVID-19 study period compared to the same months in 2019. Again, we chose to filter English-only Twitter posts given both algorithmic limitations of our methodologies and lower prevalence of non-English data (particularly in the US context). However, future work can extend our methodologies to conduct analyses in other languages to draw richer insights.

Despite the strengths of Twitter as a data source that provides us unobtrusive access to large-scale, unstructured, and naturalistic data of people's candid self-disclosure and that it has been a valuable source to study disaster and crisis response [109], we acknowledge that this data inherently has many biases such as self-selection and representation [110]. We can only study those who self-select to express on Twitter. Pew Internet

Center surveys reported that social media platforms are underrepresentative of minorities, although Twitter is an exception, which overrepresents minorities such as Blacks, Hispanics, and women [20]. There is already a digital divide in terms of social media use where the population is skewed toward young adults and white-collar workers. Further, technology and social media could be a luxury to marginalized and underprivileged populations, and any sort of technology-driven support and assistance will disproportionately affect different individuals based on technology use [88]. Similarly, a single platform cannot provide a complete picture; different platforms (eg, Facebook, Reddit, Twitter, and instant messaging services) have unique design strengths and weaknesses both in terms of their affordances as well as who uses them. Therefore, as highlighted in a recent article by Chunara and Cook [23], public health surveillance (including that for COVID-19) can account for several factors such as the "population at risk" in epidemiology and demographic disparities in the use and behavioral expressions on social media.

We understand that our study is observational and, as any other observational study, does not measure "true causality." Watts [111] noted the impossibility to test all explanations and confounders simultaneously. However, by including and comparing against control data, we minimized geo-temporal and seasonal confounds, thereby enabling us to provide stronger evidence and insights than purely correlational analyses regarding the effects of the COVID-19 pandemic on people's mental health. We also note that support expressions in our study can not only include support-seeking but also support-providing expressions. This has also enabled us to observe how solidarity and sense of belonging proliferated during the COVID-19 crisis. Future work can build separate high precision classifiers for each kind of expression to disentangle the prevalence of seeking and providing expressions during the crisis.

Further, although we did have data beyond May 24, 2020, we decided to exclude those to keep our focus on the effects on social media expressions due to COVID-19 and minimize those that followed the death of George Floyd on May 25, 2020, in the light of the Black Lives Matter protests throughout the United States [112]. We also are aware that, with the continuing nature of the pandemic, our conclusions are restricted to the mental health and support-seeking concerns expressed during a finite study period. Events since the end of the study period underscore the dynamic nature of these events, as different parts of the United States are heavily affected, while others are recovering, and some remain relatively spared. It will be important to extend this work temporally; increase the size of future samples; and, whenever possible, add geospatial specificity to future analyses. The latter will be especially important for potential supportive interventions locally if one has the resources and the ability to assemble recurring, near-real-time local "snapshots" as a basis for community-focused preventive interventions. Further, our analyses can be extended to retrospectively examine the aftereffects of particular global and local events, such as policy changes, related to the pandemic.

Conclusion

Our study, like those of others studying other major events, further reinforces the potential utility of accessing and analyzing social media data in near-real-time to *take the temperature* of communities. This will require a more focused and robust collection of locally targeted information to build samples that are sufficiently large to produce reliably representative data sets to be useful for public health interventions. Further work is now needed to track mental health–related expressions and those reflecting needs for support throughout the pandemic, which has seen dynamic changes associated with disease spread to areas that had been less affected during the early months of the outbreak. This geo-specific research may further enhance our understanding of the causal connections between COVID-19 spread and waves of expressed distress. Having the ability to present locally pertinent, contemporaneous analyses offers the opportunity for local public health and mental health providers as well as political leaders to develop and deploy targeted support services in a timely fashion.

Conflicts of Interest

JT receives unrelated research support from Otsuka.

References

1. Hays JN. Epidemics and Pandemics: Their Impacts on Human History. Santa Barbara, CA: ABC-CLIO; 2005.
2. Boscarino JA. Community disasters, psychological trauma, and crisis intervention. Int J Emerg Ment Health 2015;17(1):369-371 [FREE Full text] [Medline: 25983663]
3. Lu S. An epidemic of fear. American Psychological Assocation. 2015. URL: https://www.apa.org/monitor/2015/03/fear [accessed 2020-07-25]
4. Morganstein JC, Fullerton CS, Ursano RJ, Donato D, Holloway HC. 18 - pandemics: health care emergencies. In: Ursano RJ, Fullerton CS, Weisaeth L, Raphael B, editors. Textbook of Disaster Psychiatry. Cambridge, England: Cambridge University Press; 2017:270-284.
5. COVID-19 coronavirus pandemic. Worldometers. URL: https://www.worldometers.info/coronavirus/ [accessed 2020-07-25]
6. Basu T. The coronavirus pandemic is a game changer for mental health care. MIT Technology Review. 2020. URL: https://www.technologyreview.com/2020/03/20/905184/coronavirus-online-therapy-mental-health-app-teletherapy/ [accessed 2020-07-25]
7. Resnick B. A third of Americans report anxiety or depression symptoms during the pandemic. Vox. 2020. URL: https://www.vox.com/science-and-health/2020/5/29/21274495/pandemic-cdc-mental-health [accessed 2020-07-25]
8. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological distress and loneliness reported by US adults in 2018 and April 2020. JAMA 2020 Jul 07;324(1):93-94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
9. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
10. World Health Organization. Operational considerations for case management of COVID-19 in health facility and community. Interim guidance. Pediatr Med Rodz 2020 Jul 14;16(1):27-32. [doi: 10.15557/pimr.2020.0004]
11. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
12. Saxena SK, editor. Coronavirus Disease 2019 (COVID-19) Epidemiology, Pathogenesis, Diagnosis, and Therapeutics. Singapore: Springer; 2020.
13. Miller G. Social distancing prevents infections, but it can have unintended consequences. Science 2020 Mar 17:1. [doi: 10.1126/science.abb7506]
14. Varatharaj A, Thomas N, Ellul MA, Davies NWS, Pollak TA, Tenorio EL, CoroNerve Study Group. Neurological and neuropsychiatric complications of COVID-19 in 153 patients: a UK-wide surveillance study. Lancet Psychiatry 2020 Oct;7(10):875-882 [FREE Full text] [doi: 10.1016/S2215-0366(20)30287-X] [Medline: 32593341]
15. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
16. Fiorillo A, Gorwood P. The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. Eur Psychiatry 2020 Apr 01;63(1):e32 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.35] [Medline: 32234102]
17. Heymann DL, Shindo N, WHO Scientific and Technical Advisory Group for Infectious Hazards. COVID-19: what is next for public health? Lancet 2020 Feb 22;395(10224):542-545 [FREE Full text] [doi: 10.1016/S0140-6736(20)30374-3] [Medline: 32061313]

18. Weinstock CP. Ripple effects of COVID-19 strain mental health systems. US News. 2020. URL: https://www.usnews.com/news/healthiest-communities/articles/2020-06-04/coronavirus-ripple-effects-strain-mental-health-systems [accessed 2020-07-25]

19. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43. [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]

20. Social media fact sheet. Pew Internet Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-07-25]

21. Andalibi N, Haimson OL, De Choudhury M, Forte A. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 2016; San Jose, CA. [doi: 10.1145/2858036.2858096]

22. Qiu L, Lin H, Leung AK, Tov W. Putting their best foot forward: emotional disclosure on Facebook. Cyberpsychol Behav Soc Netw 2012 Oct;15(10):569-572. [doi: 10.1089/cyber.2012.0200] [Medline: 22924675]

23. Chunara R, Cook SH. Using digital data to protect and promote the most vulnerable in the fight against COVID-19. Front Public Health 2020;8:296. [doi: 10.3389/fpubh.2020.00296] [Medline: 32596201]

24. Guntuku SC, Schwartz HA, Kashyap A, Gaulton JS, Stokes DC, Asch DA, et al. Variability in language used on social media prior to hospital visits. Sci Rep 2020 Mar 12;10(1):4346. [doi: 10.1038/s41598-020-60750-8] [Medline: 32165648]

25. Saha K, Sugar B, Torous J, Abrahao B, Kıcıman E, De Choudhury M. A social media study on the effects of psychiatric medication use. Proc Int AAAI Conf Weblogs Soc Media 2019 Jun 07;13:440-451 [FREE Full text] [Medline: 32280562]

26. Saha K, Kim SC, Reddy MD, Carter AJ, Sharma E, Haimson OL, et al. The language of LGBTQ+ minority stress experiences on social media. Proc ACM Hum Comput Interact 2019 Nov;3(CSCW):1-22. [doi: 10.1145/3361108] [Medline: 32935081]

27. Ernala SK, Labetoulle T, Bane F, Birnbaum ML, Rizvi AF, Kane JM, et al. Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. 2018 Presented at: Twelfth International AAAI Conference on Web and Social Media; June 2018; Palo Alto, CA.

28. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: The 7th International AAAI Conference on Weblogs and Social Media; July 2013; Cambridge, MA.

29. Saha K, Weber I, Birnbaum ML, De Choudhury M. Characterizing awareness of schizophrenia among Facebook users by leveraging Facebook advertisement estimates. J Med Internet Res 2017 May 08;19(5):e156 [FREE Full text] [doi: 10.2196/jmir.6815] [Medline: 28483739]

30. Dredze M. How social media will change public health. IEEE Intell Syst 2012 Jul;27(4):81-84 [FREE Full text] [doi: 10.1109/MIS.2012.76]

31. Paul MJ, Dredze M. You are what you tweet: analyzing Twitter for public health. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 2011; Barcelona, Spain.

32. Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. 2012 Presented at: Twenty-Sixth AAAI Conference on Artificial Intelligence; July 2012; Toronto, ON.

33. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]

34. Saha K, Chan L, De Barbaro K, Abowd GD, De Choudhury M. Inferring mood instability on social media by leveraging ecological momentary assessments. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-27. [doi: 10.1145/3130960]

35. Coppersmith GA, Harman CT, Dredze MH. Measuring post traumatic stress disorder in Twitter. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; 2014; Ann Arbor, MI.

36. Kiciman E, Counts S, Gasser M. Using longitudinal social media analysis to understand the effects of early college alcohol use. 2018 Presented at: 12th International AAAI Conference on Web and Social Media; June 2018; Palo Alto, CA.

37. Liu J, Weitzman E, Chunara R. Assessing behavioral stages from social media data. CSCW Conf Comput Support Coop Work 2017;2017:1320-1333 [FREE Full text] [doi: 10.1145/2998181.2998336] [Medline: 29034371]

38. Sharma E, De Choudhury M. Mental health support and its relationship to linguistic accommodation in online communities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 2018; Montreal, QC. [doi: 10.1145/3173574.3174215]

39. Saha K, Sharma A. Causal factors of effective psychosocial outcomes in online mental health communities. Proc Int AAAI Conference Web Soc Media 2020;14(1):590-601.

40. Filter realtime tweets. Twitter Developer. URL: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/guides/basic-stream-parameters [accessed 2020-08-17]

41. Pfeffer J, Mayer K, Morstatter F. Tampering with Twitter's sample API. EPJ Data Sci 2018 Dec 19;7(1):1. [doi: 10.1140/epjds/s13688-018-0178-0]

42. Morstatter F, Pfeffer J, Liu H. When is it biased?: assessing the representativeness of twitter's streaming API. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14 Companion; April 2014; Seoul, Korea. [doi: 10.1145/2567948.2576952]

43. Boyd D, Crawford K. Critical questions for big data. Inf Commun Soc 2012 Jun;15(5):662-679. [doi: 10.1080/1369118X.2012.678878]

44. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. "Right time, right place" health communication on Twitter: value and accuracy of location information. J Med Internet Res 2012 Nov 15;14(6):e156 [FREE Full text] [doi: 10.2196/jmir.2121] [Medline: 23154246]

45. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. J Med Internet Res 2014 May 27;16(5):e137 [FREE Full text] [doi: 10.2196/jmir.3298] [Medline: 24867458]

46. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/ [accessed 2020-05-09]

47. Timeline of the COVID-19 pandemic in the United States. Wikipedia. URL: https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_the_United_States [accessed 2020-07-25]

48. Saha K, De Choudhury M. Modeling stress with social media around incidents of gun violence on college campuses. Proc ACM Hum-Comput Interact 2017 Dec 06;1(CSCW):1-27. [doi: 10.1145/3134727]

49. Iqbal M. Twitter revenue and usage statistics (2020). Business of Apps. 2020. URL: https://www.businessofapps.com/data/twitter-statistics/#2 [accessed 2020-08-17]

50. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2098-2110 [FREE Full text] [doi: 10.1145/2858036.2858207] [Medline: 29082385]

51. Guntuku SC, Ramsay JR, Merchant RM, Ungar LH. Language of ADHD in adults on social media. J Atten Disord 2019 Oct;23(12):1475-1485. [doi: 10.1177/1087054717738083] [Medline: 29115168]

52. Saha K, Weber I, De Choudhury M. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. Proc Int AAAI Conf Weblogs Soc Media 2018 Jun;2018:320-329 [FREE Full text] [Medline: 30505628]

53. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, WHO Strategic and Technical Advisory Group for Infectious Hazards. COVID-19: towards controlling of a pandemic. Lancet 2020 Mar 28;395(10229):1015-1018 [FREE Full text] [doi: 10.1016/S0140-6736(20)30673-5] [Medline: 32197103]

54. Taha S, Matheson K, Cronin T, Anisman H. Intolerance of uncertainty, appraisals, coping, and anxiety: the case of the 2009 H1N1 pandemic. Br J Health Psychol 2014 Sep;19(3):592-605. [doi: 10.1111/bjhp.12058] [Medline: 23834735]

55. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.

56. Bagroy S, Kumaraguru P, De Choudhury M. A social media based index of mental well-being in college campuses. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May;2017:1634-1646 [FREE Full text] [doi: 10.1145/3025453.3025909] [Medline: 28840202]

57. Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]

58. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]

59. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; May 2014; Ann Arbor, MI.

60. Rains SA, Young V. A meta-analysis of research on formal computer-mediated support groups: examining group characteristics and health outcomes. Hum Commun Res 2009;35(3):309-336. [doi: 10.1111/j.1468-2958.2009.01353.x]

61. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Proc Int AAAI Conf Weblogs Soc Media 2017 May;2017:32-41 [FREE Full text] [Medline: 28840079]

62. Oh HJ, Lauckner C, Boehmer J, Fewins-Bliss R, Li K. Facebooking for health: an examination into the solicitation and effects of health-related social support on social networking sites. Comput Hum Behav 2013 Sep;29(5):2072-2080. [doi: 10.1016/j.chb.2013.04.017]

63. Luo C, Li Y, Chen A, Tang Y. What triggers online help-seeking retransmission during the COVID-19 period? Empirical evidence from Chinese social media. PLoS One 2020;15(11):e0241465 [FREE Full text] [doi: 10.1371/journal.pone.0241465] [Medline: 33141860]

64. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 1992;19(2):154-174. [doi: 10.1177/009365092019002002]

65. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304. [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]

66. Andalibi N, Haimson OL, Choudhury MD, Forte A. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. ACM Trans Comput-Hum Interact 2018 Oct 17;25(5):1-35. [doi: 10.1145/3234942]

67. Wang YC, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 2012; Seattle, WA. [doi: 10.1145/2145204.2145329]

68. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]

69. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]

70. Eisenstein J, Ahmed A, Xing EP. Sparse additive generative models of text. 2011 Presented at: 28th International Conference on International Conference on Machine Learning; June 2011; Bellevue, WA.

71. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Taylor & Francis Group; 2013.

72. Rankin CH, Abrams T, Barry RJ, Bhatnagar S, Clayton DF, Colombo J, et al. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiol Learn Mem 2009 Sep;92(2):135-138 [FREE Full text] [doi: 10.1016/j.nlm.2008.09.012] [Medline: 18854219]

73. De Choudhury M, Monroy-Hernández A, Mark G. "Narco" emotions: affect and desensitization in social media during the mexican drug war. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 2014; Toronto, ON. [doi: 10.1145/2556288.2557197]

74. Medina J, Mele C, Murphy H. One dead in synagogue shooting near San Diego; officials call it hate crime. The New York Times. 2019 Apr. URL: https://www.nytimes.com/2019/04/27/us/poway-synagogue-shooting.html [accessed 2020-07-25]

75. Anthony S. Fauci, M.D. National Institute of Allergy and Infectious Diseases. URL: https://www.niaid.nih.gov/about/anthony-s-fauci-md-bio [accessed 2020-07-25]

76. Colson P, Rolain J, Lagier J, Brouqui P, Raoult D. Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. Int J Antimicrob Agents 2020 Apr;55(4):105932 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105932] [Medline: 32145363]

77. Mental health and psychosocial considerations during the COVID-19 outbreak. World Health Organization. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/331490/WHO-2019-nCoV-MentalHealth-2020.1-eng.pdf [accessed 2020-07-20]

78. Broom D. A pandemic of solidarity? This is how people are supporting one another as coronavirus spreads. World Economic Forum. 2020. URL: https://www.weforum.org/agenda/2020/03/covid-19-coronavirus-solidarity-help-pandemic/ [accessed 2020-08-18]

79. Free C. People across the country are delivering groceries free. It's 'solidarity, not charity.'. The Washington Post. 2020. URL: https://www.washingtonpost.com/lifestyle/2020/04/27/people-across-country-are-delivering-groceries-free-its-solidarity-not-charity/ [accessed 2020-07-25]

80. Reger MA, Stanley IH, Joiner TE. Suicide mortality and coronavirus disease 2019-a perfect storm? JAMA Psychiatry 2020 Apr 10:1. [doi: 10.1001/jamapsychiatry.2020.1060] [Medline: 32275300]

81. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004 Oct;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]

82. González-Sanguino C, Ausín B, Castellanos M, Saiz J, López-Gómez A, Ugidos C, et al. Mental health consequences during the initial stage of the 2020 coronavirus pandemic (COVID-19) in Spain. Brain Behav Immun 2020 Jul;87:172-176 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.040] [Medline: 32405150]

83. Labouliere CD, Vasan P, Kramer A, Brown G, Green K, Rahman M, et al. "Zero Suicide" - a model for reducing suicide in United States behavioral healthcare. Suicidologi 2018;23(1):22-30 [FREE Full text] [Medline: 29970972]

84. Florida R. The lasting normal for the post-pandemic city. Bloomberg. 2020. URL: https://www.bloomberg.com/news/features/2020-06-25/the-new-normal-after-the-coronavirus-pandemic [accessed 2020-07-25]

85. Asonye C. There's nothing new about the 'new normal'. Here's why. World Economic Forum. 2020. URL: https://www.weforum.org/agenda/2020/06/theres-nothing-new-about-this-new-normal-heres-why/ [accessed 2020-07-25]

86. Rudd BN, Beidas RS. Digital mental health: the answer to the global mental health crisis? JMIR Ment Health 2020 Jun 02;7(6):e18472 [FREE Full text] [doi: 10.2196/18472] [Medline: 32484445]

87. Yoo DW, Birnbaum ML, Van Meter AR, Ali AF, Arenare E, Abowd GD, et al. Designing a clinician-facing tool for using insights from patients' social media activity: iterative co-design approach. JMIR Ment Health 2020 Aug 12;7(8):e16969 [FREE Full text] [doi: 10.2196/16969] [Medline: 32784180]

88. Pendse SR, Lalani FM, De Choudhury M, Sharma A, Kumar N. "Like Shock Absorbers": understanding the human infrastructures of technology-mediated mental health support. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 2020; Honolulu, HI. [doi: 10.1145/3313831.3376465]

89. Saha K, Bayraktaroglu AE, Campbell AT, Chawla NV, De Choudhury M, D'Mello SK, et al. Social media as a passive sensor in longitudinal studies of human behavior and wellbeing. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 2019; Glasgow, Scotland. [doi: 10.1145/3290607.3299065]

90. Cunha BA. Influenza: historical aspects of epidemics and pandemics. Infect Dis Clin North Am 2004 Mar;18(1):141-155. [doi: 10.1016/S0891-5520(03)00095-3] [Medline: 15081510]

91. Aassve A, Alfani G, Gandolfi F, Le Moglie M. Epidemics and trust: the case of the spanish flu. Working Papers 661, Innocenzo Gasparini Institute for Economic Research 2020:1.

92. Phillips H. The recent wave of 'Spanish' flu historiography. Social Hist Med 2014 Sep 01;27(4):789-808. [doi: 10.1093/shm/hku066]

93. Huremović D. Psychiatry of Pandemics: a Mental Health Response to Infection Outbreak. Cham, Switzerland: Springer; 2019.

94. Ornell F, Schuch JB, Sordi AO, Kessler FHP. "Pandemic fear" and COVID-19: mental health burden and strategies. Braz J Psychiatry 2020;42(3):232-235 [FREE Full text] [doi: 10.1590/1516-4446-2020-0008] [Medline: 32267343]

95. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 1995 May;117(3):497-529. [Medline: 7777651]

96. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

97. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: a systematic narrative review. Child Youth Services Rev 2014 Jun;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]

98. Berryman C, Ferguson CJ, Negy C. Social media use and mental health among young adults. Psychiatr Q 2018 Jun;89(2):307-314. [doi: 10.1007/s11126-017-9535-6] [Medline: 29090428]

99. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]

100. Saha K, Torous J, Ernala SK, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med 2019 Nov 25;9(6):1197-1207 [FREE Full text] [doi: 10.1093/tbm/ibz028] [Medline: 30834942]

101. Nelson BW, Pettitt A, Flannery JE, Allen NB. Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. PLoS One 2020;15(11):e0241990 [FREE Full text] [doi: 10.1371/journal.pone.0241990] [Medline: 33175882]

102. Lublin N. Notes on coronavirus: how is America feeling? Part 7. Crisis Text Line. 2020. URL: https://www.crisistextline.org/mental-health/notes-on-coronavirus-how-is-america-feeling-part-7/ [accessed 2020-07-25]

103. Miltimore J. 'A Year's Worth of Suicide Attempts in Four Weeks': the unintended consequences of COVID-19 lockdowns. Foundation for Economic Education. 2020. URL: https://fee.org/articles/a-years-worth-of-suicide-attempts-in-four-weeks-the-unintended-consequences-of-covid-19-lockdowns/ [accessed 2020-07-25]

104. Seaman J. Suicides in Colorado dropped 40% during first 2 months of coronavirus pandemic but calls to crisis line spiked. The Denver Post. 2020. URL: https://www.denverpost.com/2020/05/23/colorado-suicides-dropped-coronavirus-pandemic-calls-to-crisis-line-spiked/ [accessed 2020-07-25]

105. Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O'Malley AJ, et al. Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. JMIR Preprints. Preprint posted online on May 26, 2020. [doi: 10.2196/preprints.19347]

106. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]

107. Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public priorities and concerns regarding COVID-19 in an online discussion forum: longitudinal topic modeling. J Gen Intern Med 2020 Jul;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]

108. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking mental health and symptom mentions on Twitter during COVID-19. J Gen Intern Med 2020 Sep;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]

109. Imran M, Castillo C, Diaz F, Vieweg S. Processing social media messages in mass emergency: survey summary. In: Companion Proceedings of the Web Conference 2018. 2018 Presented at: WWW '18; April 2018; Lyon, France. [doi: 10.1145/3184558.3186242]

110. Olteanu A, Castillo C, Diaz F, Kıcıman E. Social data: biases, methodological pitfalls, and ethical boundaries. Front Big Data 2019 Jul 11;2:13. [doi: 10.3389/fdata.2019.00013]

111. Watts DJ. Common sense and sociological explanations. AJS 2014 Sep;120(2):313-351. [doi: 10.1086/678271] [Medline: 25811066]

112. George Floyd: what happened in the final moments of his life. BBC News. 2020. URL: https://www.bbc.com/news/world-us-canada-52861726 [accessed 2020-07-25]

Abbreviations

API: application programming interface
ES: emotional support
IS: informational support
LLR: log-likelihood ratio
SAGE: Sparse Additive Generative Model
SVM: support vector machine

XSL•FO
RenderX

<u>Original Paper</u>

Digital Health Technologies Respond to the COVID-19 Pandemic In a Tertiary Hospital in China: Development and Usability Study

Wanmin Lian[1*], MSc; Li Wen[1*], MSc; Qiru Zhou[2], BA; Weijie Zhu[1], BA; Wenzhou Duan[1], MSc; Xiongzhi Xiao[1], MSc; Florence Mhungu[3], PhD; Wenchen Huang[4], MSc; Chongchong Li[5], BA; Weibin Cheng[6], MSc; Junzhang Tian[6], MD

[1]Information Department, Guangdong Second Provincial General Hospital, Guangzhou, China

[2]Internet Hospital, Guangdong Second Provincial General Hospital, Guangzhou, China

[3]Department of Toxicology, School of Public Health, Southern Medical University, Guangzhou, China

[4]Dott Medical Co Ltd, Shenzhen, China

[5]Beijing Rxthinking Technology Co Ltd, Beijing, China

[6]Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Weibin Cheng, MSc
Institute for Healthcare Artificial Intelligence Application
Guangdong Second Provincial General Hospital
No 466 Xingangzhong Road
Haizhu District
Guangzhou, 510317
China
Phone: 86 13929587509
Email: chwb817@gmail.com

Abstract

Background: The outbreak of COVID-19 has caused a continuing global pandemic. Hospitals are integral to the control and prevention of COVID-19; however, they are facing numerous challenges during the epidemic.

Objective: Our study aimed to introduce the practical experience of the design and implementation of a web-based COVID-19 service platform at a tertiary hospital in China as well as the preliminary results of the implementation.

Methods: The web-based COVID-19 service platform was deployed within the health care system of the Guangdong Second Provincial General Hospital and Internet Hospital; the function of the platform was to provide web-based medical services for both members of the public and lay health care workers. The focal functions of this system included automated COVID-19 screening, related symptom monitoring, web-based consultation, and psychological support; it also served as a COVID-19 knowledge hub. The design and process of each function are introduced. The usage data for the platform service were collected and are represented by three periods: the pre-epidemic period (December 22, 2019, to January 22, 2020, 32 days), the controlled period (January 23 to March 31, 2020, 69 days), and the postepidemic period (April 1 to June 30, 2020, 91 days).

Results: By the end of June 2020, 96,642 people had used the automated COVID-19 screening and symptom monitoring systems 161,884 and 7,795,194 times, respectively. The number of general web-based consultation services per day increased from 30 visits in the pre-epidemic period to 122 visits during the controlled period, then dropped to 73 visits in the postepidemic period. The psychological counseling program served 636 clients during the epidemic period. For people who used the automated COVID-19 screening service, 160,916 (99.40%) of the total users were classified in the no risk category. 464 (0.29%) of the people were categorized as medium to high risk, and 12 people (0.01%) were recommended for further COVID-19 testing and treatment. Among the 96,642 individuals who used the COVID-19 related symptoms monitoring service, 6696 (6.93%) were symptomatic at some point during the monitoring period. Fever was the most frequently reported symptom, with 2684/6696 symptomatic people (40.1%) having had this symptom. Cough and sore throat were also relatively frequently reported by the 6696 symptomatic users (1657 people, 24.7%, and 1622 people, 24.2%, respectively).

Conclusions: The web-based COVID-19 service platform implemented at a tertiary hospital in China is exhibited to be a role model for using digital health technologies to respond to the COVID-19 pandemic. The digital solutions of automated COVID-19

XSL·FO
RenderX

screening, daily symptom monitoring, web-based care, and knowledge propagation have plausible acceptability and feasibility for complementing offline hospital services and facilitating disease control and prevention.

KEYWORDS

Internet hospital; COVID-19; automated screening; symptom; monitoring; web-based consultation; psychological support; emergency; digital health; hospital; China; screening

Introduction

The outbreak of COVID-19 has caused an ongoing global pandemic that is presently affecting over 37 million people worldwide. Hospitals have been at the center of the COVID-19 control and prevention effort while facing many challenges during the epidemic. The surging demands of health care for COVID-19 screening and treatment have overwhelmed the medical system [1,2], and the lack of proper personal protective equipment for medical staff is causing nosocomial infection concerns [3,4]. Additionally, maintaining routine care services such as chronic condition care and emergency outpatient visits while suspending general outpatient visits during the epidemic has placed stress on hospitals [5-7].

The Guangdong Second Provincial General Hospital (GD2H) is a large-scale tertiary hospital located in Guangzhou, China, and it is renowned for its emergency medical rescue and smart hospital services. GD2H established the first internet hospital in China in October 2014, and it is a pioneering center in the exploration of smart hospitals [8]. In 2019, the internet hospital was upgraded with a new application equipped with over 20 digital health technologies, including artificial intelligence (AI) physician services, distance electrocardiogram diagnosis, a prescription circulation platform, and a medical imaging diagnosis system. These digital technologies help shift tasks from the hospital to the community. Based on the internet hospital, GD2H established a tiered health care delivery system that provides professional medical care support for lay health care workers in 2377 poor villages in Guangdong, China. In the application of 5G technology, GD2H took the lead in establishing the 5G distance surgery practice in Guangdong Province. Meanwhile, as the first provincial emergency hospital in China and the seventh World Health Organization Emergency Medical Team, GD2H was one of the major designated COVID-19 treatment centers in response to the COVID-19 outbreak in Guangdong, China.

Digital health solutions, including internet hospitals, have been reported to facilitate epidemic control measures such as contact tracing and prehospital triaging while providing web-based medical care [9-14]. There has been a surge in the establishment of internet hospitals during the COVID-19 pandemic. A total of 213 new internet hospitals (compared to 362 by the end of 2019) were established between January and June 2020 in China. Over 47,000 physicians voluntarily provide medical care services on Haodf.com (a private internet hospital platform) [15]. However, less information has been documented concerning the implementation perspective of these technologies in hospitals [16]. In this paper, we introduce the practical experience of design and implementation as well as the preliminary results of an internet hospital–based web-based COVID-19 service platform; its functions include automated COVID-19 screening, monitoring of related symptoms, and web-based care services, and it additionally serves as a knowledge hub.

Methods

Service Framework

The web-based COVID-19 platform was deployed at the GD2H Internet Hospital [8], which provides web-based medical services for both public individuals (ie, customer clients) and lay health care workers (ie, physician clients, rural health care personnel without formal medical training). The service framework is shown in Figure 1. The upper layer shows the portal of the service, which includes WeChat (a popular social media platform in China, similar to Facebook and Twitter), a smartphone app (developed by the GD2H), a website, and target users. The main functions include automated COVID-19 screening, monitoring of related symptoms, web-based consultation, psychological support, and COVID-19 knowledge dissemination. Modern information technologies such as the Internet of Things, big data, and AI were featured in the services platform.

Figure 1. The web-based medical service framework.
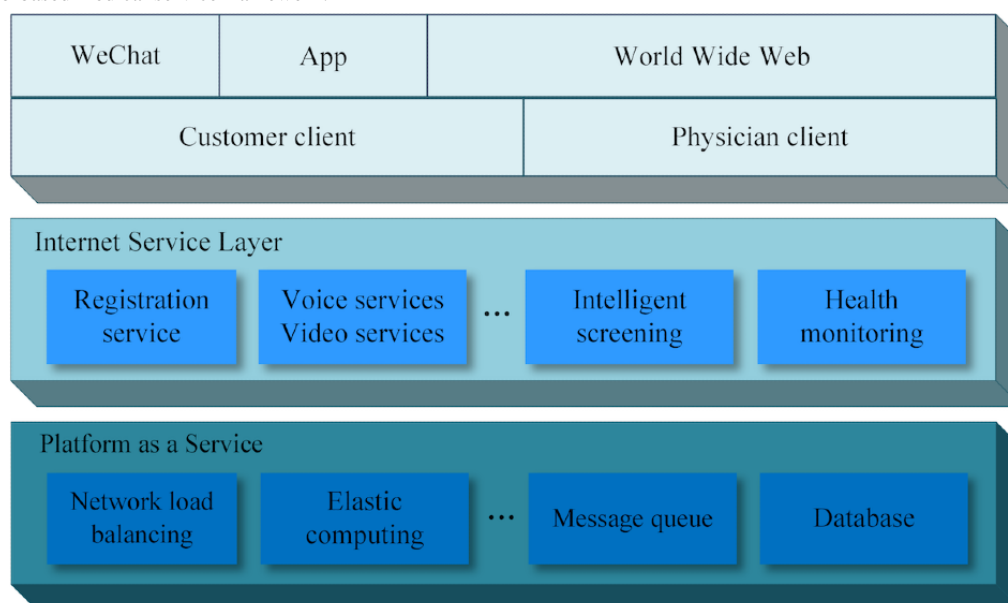
Automated COVID-19 Screening

This service functions as a forward triage strategy that enables patients to be efficiently screened before they arrive at the hospital emergency department; this strategy protects patients, clinicians, and the community from exposure. Automated screening algorithms were designed based on a decision tree that classified patients according to their symptoms, travel history, and exposure to COVID-19. Clients who met the epidemiological suspected criteria were then transferred to a web-based physician consultation for further screening and care. Inclusion of symptoms, determination of duration, and contact history were based on the COVID-19 Diagnosis and Treatment Protocol (trial version 7) issued by the National Health Commission of the People's Republic of China [17]. Based on the report, individuals were classified as no risk, low risk, medium to high risk, and high risk. The definitions of the risk categories are provided in Multimedia Appendix 1. The process of automated COVID-19 screening is shown in Figure 2.

Figure 2. Flowchart for automated COVID-19 screening.
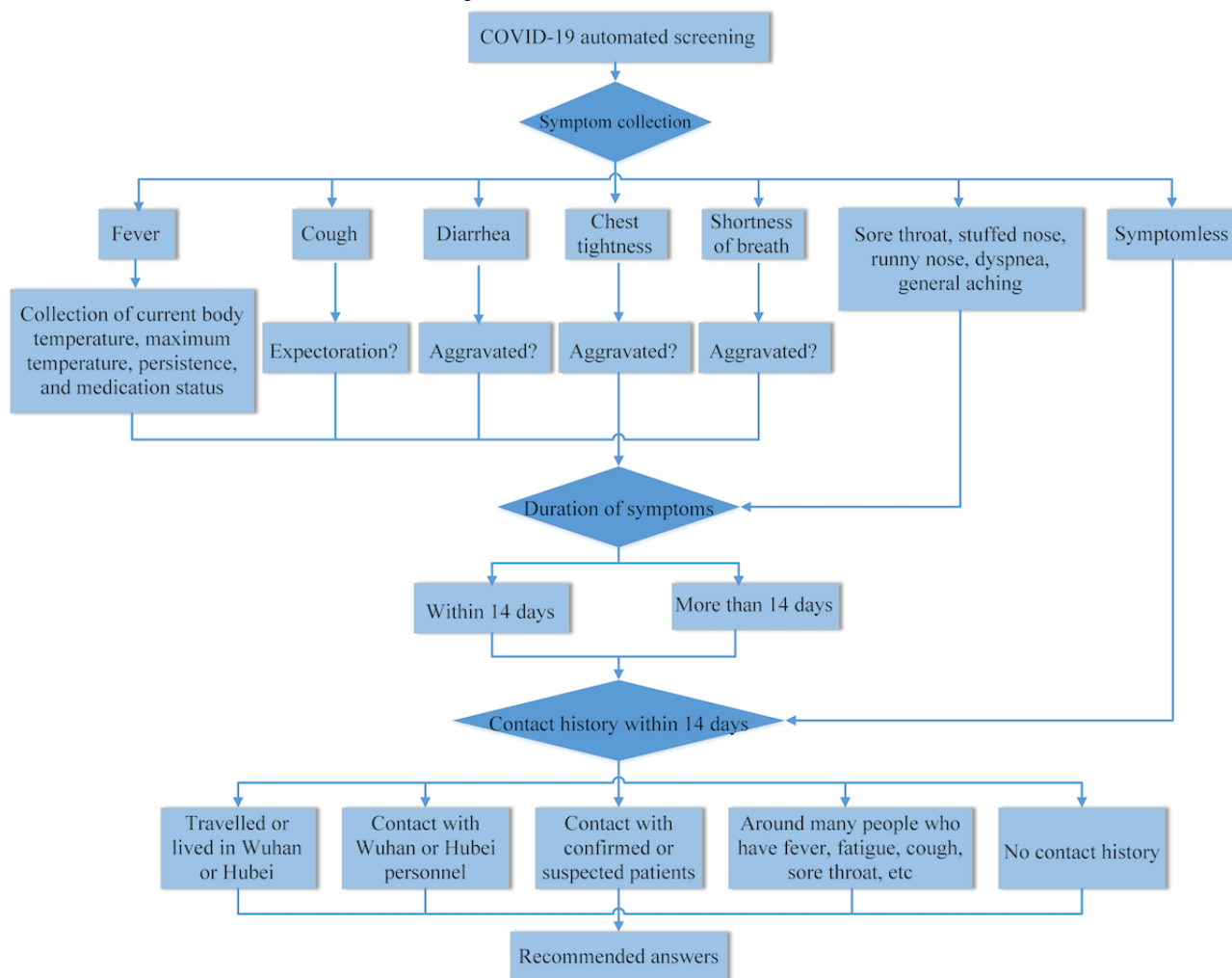
Monitoring of COVID-19–Related Symptoms

This service was designed to facilitate individuals' symptom self-management and staff health status management by organizations. The symptom monitoring protocol was based on the China Center for Disease Control and Prevention guideline for monitoring of close contacts of COVID-19 cases [18]. Typical COVID-19–related symptoms, such as fever, sore throat, and fatigue, were collected in a structured way twice daily and self-reported through the app portal. A dashboard displayed visual graphics to show the changes in symptoms for users and physicians. Physicians of the internet hospital were then alerted of any abnormal statuses in real time, and the internet hospital physician would then reach out and provide guidance to the client for risk assessment and treatment. The abnormal results that triggered the web-based consultation services included body temperature ≥38 ℃; body temperature ≥37.3 ℃ and sore throat; and body temperature ≥37.3 ℃ and fatigue.

The process of COVID-19–related symptom monitoring is shown in Figure 3.

Figure 3. Flowchart for monitoring of COVID-19–related symptoms.
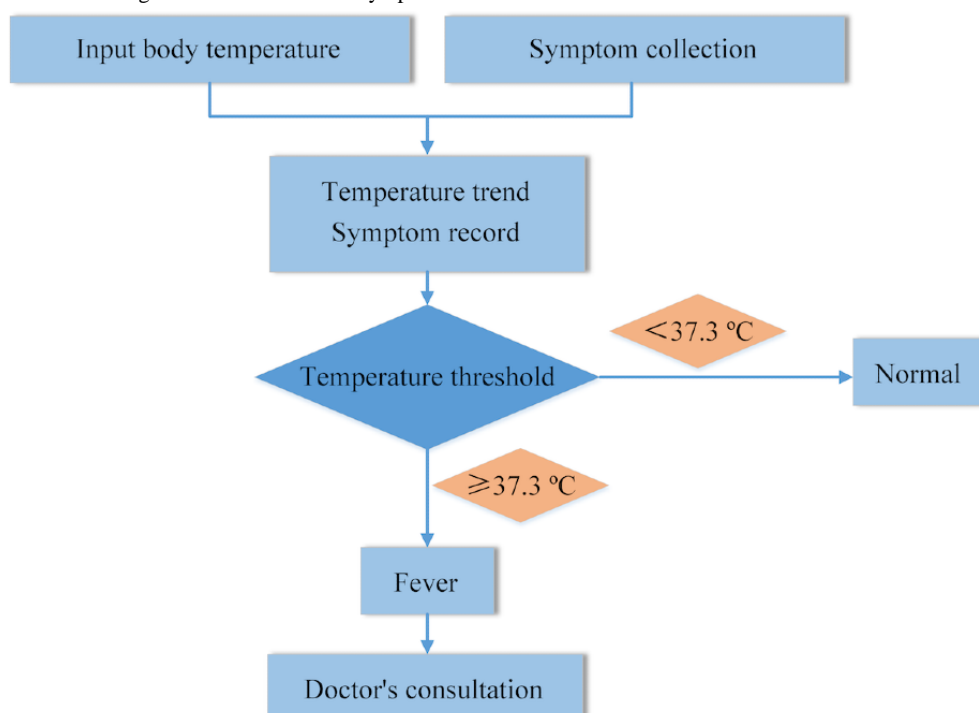
Web-Based Consultation

Web-based consultation provides free general medical services and psychological counseling for the public through WeChat and the app, which links to the GD2H Internet Hospital. This service also provides specialized medical care support for lay health care workers who encounter problems while providing medical care in their villages and for primary care centers through a desktop system. Web-based consultation services can also be triggered when the results of automated COVID-19 screening or risk monitoring are abnormal.

General medical services were provided by 30 full-time professionally qualified physicians who were registered with the GD2H internet hospital. The physicians' specialties covered internal medicine, surgery, traditional Chinese medicine, and rehabilitation. Psychological counseling services were provided by licensed psychiatrists; these services targeted not only the public and patients but also medical personnel to alleviate the psychological distress caused by the epidemic.

All physicians also received professional training on COVID-19 diagnosis and treatment. The web-based consultation service supports both text input and video consultation modes. The consultation interface enables users to upload materials including descriptions of illness, past medical history, symptoms, and test and examination results in various text, sound, and image formats; also, data uploaded by users are encrypted to protect privacy. Physicians prescribed authenticated digital prescriptions on the web, which were connected to a third-party drug distribution agency that provides home delivery service of drugs. The process of web-based COVID-19 consultation is shown in Figure 4.

Figure 4. Flowchart of web-based consultation. ECG: electrocardiogram.
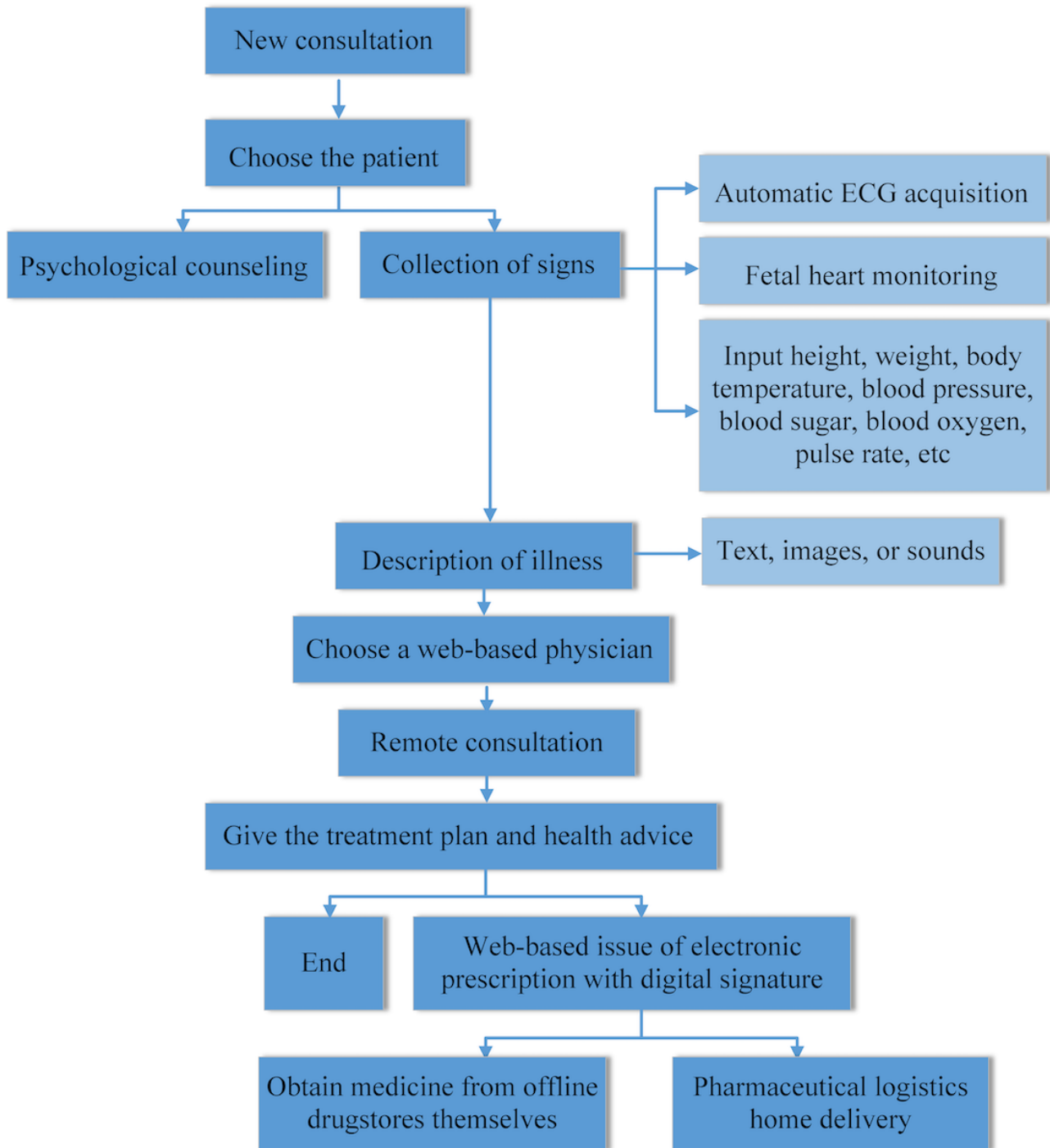
COVID-19 Knowledge Hub

The COVID-19 knowledge hub provides a COVID-19 encyclopedia to educate the public and help them understand the COVID-19 epidemic in a timely and comprehensive manner. COVID-19–related policies and prevention measures, treatment, and other information is appropriately updated. When new COVID-19 prevention and control information, such as updated guidelines and expert consensus, is officially released by the National Health Commission and the Chinese Society of Nutrition, it is pushed to the platform in infographic form for clients. The COVID-19 treatment module mainly provides detailed descriptions of use specifications, dosages, adverse reactions, and contraindications of common antiviral drugs, immunopotentiators, glucocorticoids, and antibacterial drugs. Traditional Chinese medicine, including specific prescription compositions, indications, and other information, is verified through practice by the traditional medicine and sports injury rehabilitation research team of GD2H.

Data Collection and Analysis

Usage data of the platform were collected for this analysis. The automated COVID-19 screening, COVID-19–related symptom monitoring, and web-based consultation services have been in use since January 26, 2020. All data were collected up to June 30, 2020. On January 23, 2020, Wuhan was placed on official

lockdown and the Guangdong Provincial Government declared the level one public health emergency response to COVID-19. By March 31, 2020, most areas in China were classified as low risk. A week later (April 8), Wuhan lifted its COVID-19 restrictions [19], which showed that the epidemic in China had been controlled. Based on these two time points, we divided the whole period into three sub-periods: the pre-epidemic period (before January 23), controlled period (January 23 to March 31), and postepidemic period (after March 31). Daily hospital service visit data (one month) were collected to compare the hospital service volumes between the preoutbreak and postoutbreak periods. All information was deidentified before analysis. The study was approved by the ethical review board of GD2H.

Results

Hospital outpatient visits averaged 3266 per day prior to the COVID-19 outbreak, then dropped to approximately 1182 visits per day during the controlled period and slightly increased to about 2699 visits per day in the postepidemic period. Although fever clinic visits increased from 11 per day prior to the outbreak to 56 per day during the controlled period and 37 visits per day in the postepidemic period (Table 1), Figure 5 shows that the outpatient visits dropped significantly early in the outbreak and ascended slowly afterward. The trends of fever clinic visits and web-based consultations significantly increased in the early stages of the COVID-19 outbreak, then decreased to a level that was relatively low but still higher than that in the pre-epidemic period. The fever clinic visits reached a peak on March 2, 2020, 30 days after the peak of web-based consultations (Figure 6).

Table 1. Hospital outpatient service volumes and platform service usage data by functionality in the pre-epidemic, controlled, and postepidemic periods.

Functionality	Pre-epidemic period (December 22, 2019, to January 22, 2020)		Controlled period (January 23 to March 31, 2020)		Postepidemic period (April 1 to June 30, 2020)	
	Total uses	Uses per day	Total uses	Uses per day	Total uses	Uses per day
Outpatient visit	104,498	3266	81,561	1182	245,605	2699
Fever clinic visit	356	11	3886	56	3371	37
Automated COVID-19 screening	N/A[a]	N/A	93,405	1354	68,479	752
COVID-19–related symptom monitoring	N/A	N/A	2,006,178	29,075	5,775,357	63,465
General web-based consultation	960	30	8406	122	6656	73
Psychological counseling	N/A	N/A	474	7	162	2
Web-based prescription	830	26	1781	26	1898	21

[a]N/A: not applicable.

Figure 5. The trends of hospital outpatient visits from December 22, 2019, to June 30, 2020. The red lines denote the start of the controlled period and the end of the postepidemic period.
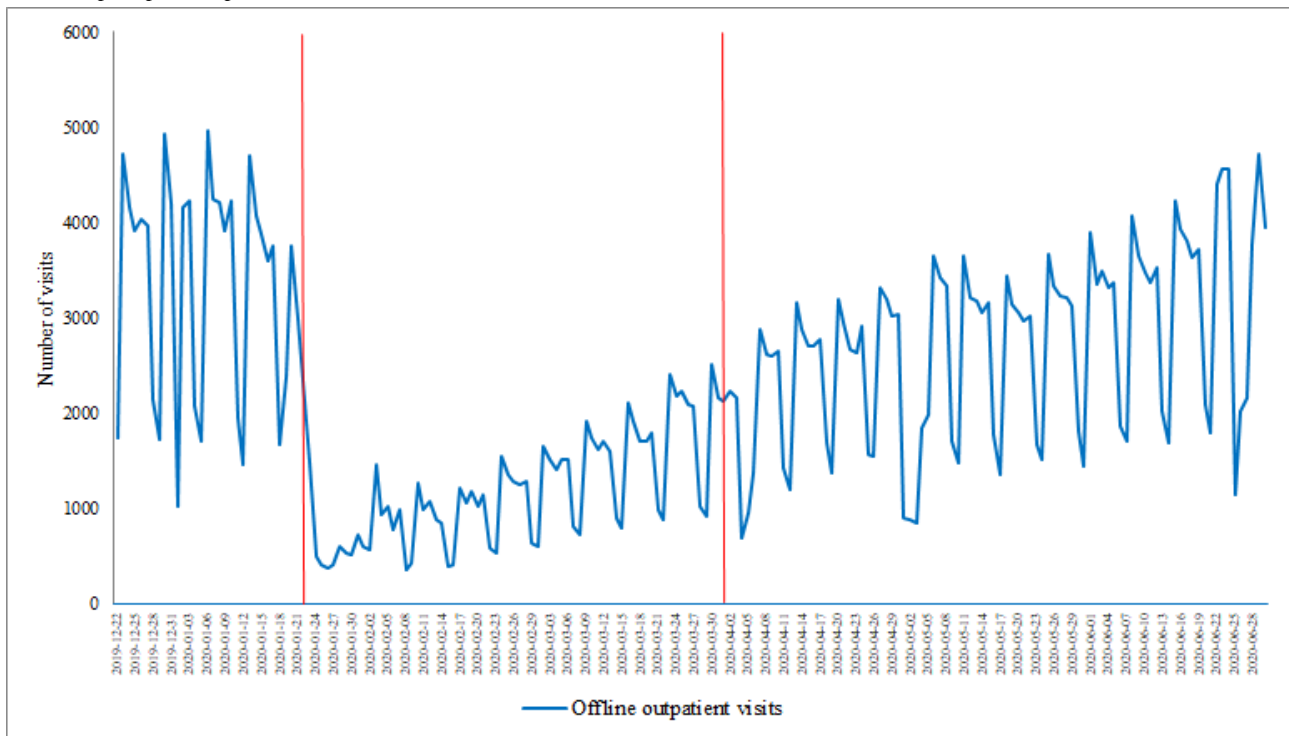
Figure 6. The trends of fever clinic visits and web-based consultation visits from December 22, 2019, to June 30, 2020. The red lines denote the start of the controlled period and the end of the postepidemic period.
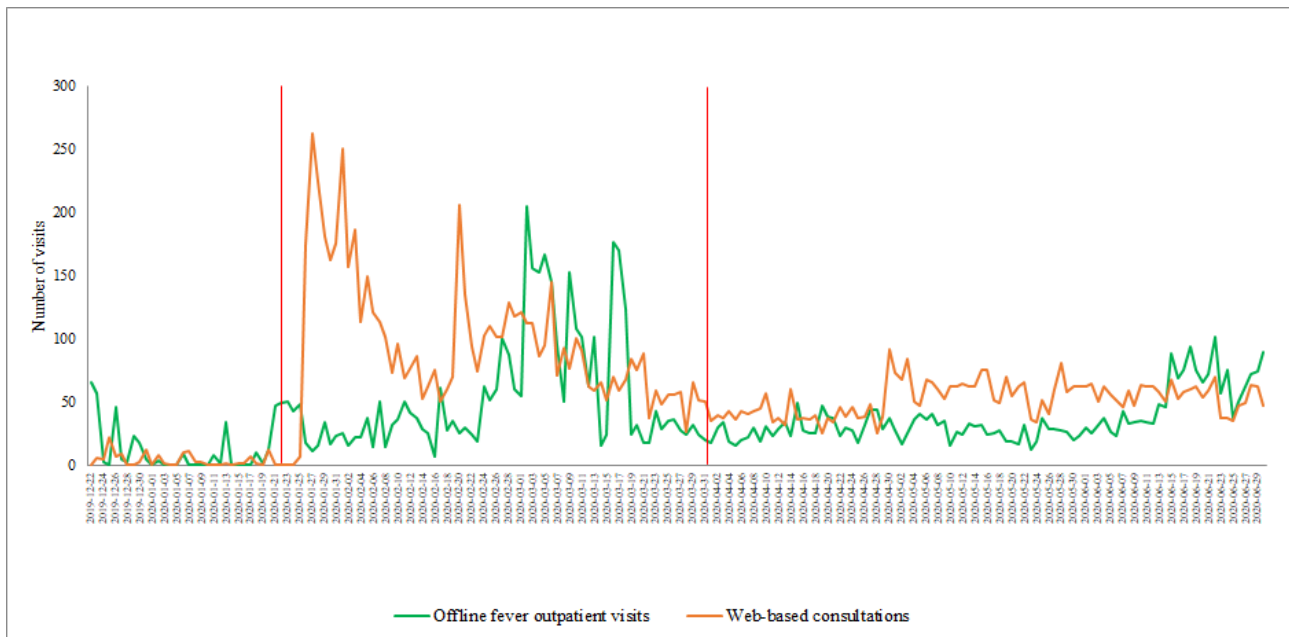
The platform service usage data are shown in Table 1. By the end of June 2020, 96,642 people had used the automated COVID-19 screening and symptom monitoring services 161,884 and 7,795,194 times, respectively. The number of general web-based consultation services per day increased from 30 visits in the pre-epidemic period to 122 visits during the controlled period, then dropped to 73 visits in the postepidemic period. The psychological counseling platform served 636 clients during the epidemic.

Of the 161,884 people who used the automated COVID-19 screening service, 160,916 users (99.40%) were classified as at no risk, 464 (0.29%) were classified as medium to high risk, and 12 (0.01%) were recommended to undergo COVID-19 testing and treatment. The results of the automated COVID-19 screening during the controlled period and postepidemic period are shown in Table 2.

Table 2. Results of automated COVID-19 screening during the controlled period and postepidemic period (N=161,884).

Result	Screens during controlled period (January 23 to March 31, 2020), n (%)	Screens during postepidemic period (April 1 to June 30, 2020), n (%)	Total screens, n (%)
Screened	93,405 (100)	68,479 (100)	161,884 (100)
No risk	92,704 (99.25)	68,212 (99.61)	160,916 (99.40)
Low risk (home observation)	435 (0.47)	57 (0.08)	492 (0.30)
Medium to high risk (quarantine)	256 (0.27)	208 (0.30)	464 (0.29)
High risk (treatment)	10 (0.01)	2 (0.01)	12 (0.01)

Among the 96,642 individuals who used the COVID-19–related symptom monitoring service, 6696 (6.9%) were symptomatic at some point during monitoring. Fever was the most frequently reported symptom; 2684/6696 symptomatic people (40.1%) had fever. Cough and sore throat were also recurrently reported by symptomatic clients (1657/6696, 24.7%, and 1622/6696, 24.3%, respectively). The results of the COVID-19–related symptom monitoring are shown in Table 3.

Table 3. Results of the COVID-19–related symptom monitoring (N=96,642).

Symptom	Controlled period (January 23 to March 31, 2020)		Postepidemic period (April 1 to June 30, 2020)	
	Persons monitored (n=46,456), n (%)	Times monitored	Persons monitored (n=50,186), n (%)	Times monitored
Asymptomatic	43,988 (94.7)	2,006,178	45,958 (91.6)	5,775,357
Symptomatic	2468 (5.3)	5673	4228 (8.4)	10,343
Fever	919 (37.2)	2666	1765 (41.7)	5219
Runny nose	241 (9.8)	382	468 (11.1)	766
Cough	623 (25.2)	1111	1036 (24.5)	1777
Sore throat	565 (22.9)	897	1057 (25.0)	1605
Fatigue	399 (16.2)	617	616 (14.6)	976

Discussion

Principal Findings

This article introduces the practical experience of the application of digital health technologies in response to the COVID-19 pandemic from the perspective of a public tertiary hospital in China. The web-based COVID-19 service platform was integrated with automated COVID-19 screening, daily symptom monitoring, web-based care services, and knowledge dissemination to achieve prehospital triage, supplement offline medical care, and facilitate disease control and prevention. Preliminary data show that this practice has good acceptability among the public and sound applicability for complementing hospital services during an emergency crisis. Our practice can serve as a structure model for hospitals to develop their own digital health services that tailor to their technology infrastructure, the disease epidemic characteristics, and the need for disease control.

This web-based COVID-19 service platform features several functionalities that respond to the pain points of disease epidemics [20]. Automated COVID-19 screening realizes prehospital triage for patients before they arrive at the hospital based on epidemiological evidence of COVID-19 using a decision tree algorithm [21]. This work substantially reduces the burden on the fever clinic service and effectively prevents nosocomial infections. Additionally, daily monitoring of COVID-19 symptoms helps administrative and health staff to manage links with web-based medical services, which guarantees a systematic flow of work. Moreover, web-based consultation provides medical services for individuals in a virtual setting, which supplements the suspension of offline medical services. Psychological counseling services provide free, professional, and systematic psychological assistance to users of web-based and offline services to prevent and alleviate the psychological distress caused by the epidemic [22]. Furthermore, this web-based service platform overcomes time and geographical limitations, which enables people to conveniently access professional care services.

The platform service usage data show that the number of web-based consultations sharply increased during the controlled period and then slightly decreased in the postepidemic period; however, this number was still far higher than that during the pre-epidemic period. In contrast, we established that the number of offline outpatient visits dropped significantly in the controlled period. These results suggest the applicability of using web-based medical services to address the challenge of maintaining medical services while reducing the likelihood of nosocomial infection.

Implications

Our work has several policy, implementation, and research implications. First, from the policy perspective, based on the experience of the COVID-19 pandemic, policy makers should

be driven toward developing a contingency plan that includes strategies of promotion and regulation of web-based medical services by defining the scopes and standards as well as the rights and responsibilities of the entities [23]. Secondly, from the implementation perspective, hospitals should plan ahead for the establishment of internet hospitals in accordance with local conditions, establish digital health technologies, and formulate emergency response measures against severe infectious disease outbreaks [24]. However, there is no "one-size-fits-all" strategy for all hospitals; the functionalities should be tailored to each hospital's needs, and the available tools should be shared within the medical consortium to achieve the highest cost-effectiveness. In addition, from the research perspective, strengthening research on web-based health services, including the scope of diagnosis and care of internet hospitals, acceptability for different subgroup populations, digital health solutions, and quality control measures, is warranted [25].

Recommendations

Although this paper has demonstrated the capabilities of this system to prevent and fight COVID-19 to a certain extent, there are several ways to further strengthen the system. First, on the premise of ensuring information security, connecting the hospital's electronic medical record database with internet hospital information can provide patients with more comprehensive and reliable medical care services. Also, customizing medical services for different groups of populations, such as web-based medical visits and home monitoring for chronic patients, can improve the efficiency of disease diagnosis and treatment and the satisfaction of patients [6]. It is also important to note that medical insurance payments for web-based medical services can further increase patients' willingness to use these services.

Conclusions

A web-based COVID-19 service platform at a tertiary hospital in China is presented as a role model for using digital health technologies to respond to the COVID-19 pandemic. The digital solutions of automated COVID-19 screening, daily symptom monitoring, web-based care, and knowledge access have commendable acceptability and feasibility for complementing offline hospital services and facilitating disease control and prevention. Future studies to evaluate the effects of relevant functions on practical applications and formulate relevant policies and measures to enhance the application of digital health technologies are of paramount importance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of the risk categories in the automated COVID-19 screening service.
[DOCX File , 14 KB - jmir_v22i11e24505_app1.docx]

References

1. Ageron F, Sarasin F, Pasquier M, Carron P. Emergency department : COVID-19 crisis and organizational aspects. Article in French. Rev Med Suisse 2020 May 06;16(692):924-929. [Medline: 32374538]
2. Elmore JG, Wang P, Kerr KF, Schriger DL, Morrison DE, Brookmeyer R, et al. Excess Patient Visits for Cough and Pulmonary Disease at a Large US Health System in the Months Prior to the COVID-19 Pandemic: Time-Series Analysis. J Med Internet Res 2020 Sep 10;22(9):e21562 [FREE Full text] [doi: 10.2196/21562] [Medline: 32791492]
3. He Y, Li W, Wang Z, Chen H, Tian L, Liu D. Nosocomial infection among patients with COVID-19: A retrospective data analysis of 918 cases from a single center in Wuhan, China. Infect Control Hosp Epidemiol 2020 Aug 13;41(8):982-983 [FREE Full text] [doi: 10.1017/ice.2020.126] [Medline: 32279676]
4. Huang H, Zhao W, Li G. Knowledge and Psychological Stress Related to COVID-19 Among Nursing Staff in a Hospital in China: Cross-Sectional Survey Study. JMIR Form Res 2020 Sep 18;4(9):e20606 [FREE Full text] [doi: 10.2196/20606] [Medline: 32640419]
5. Chiappetta S, Sharma AM, Bottino V, Stier C. COVID-19 and the role of chronic inflammation in patients with obesity. Int J Obes (Lond) 2020 Aug 14;44(8):1790-1792 [FREE Full text] [doi: 10.1038/s41366-020-0597-4] [Medline: 32409680]
6. Guarino M, Cossiga V, Fiorentino A, Pontillo G, Morisco F. Use of Telemedicine for Chronic Liver Disease at a Single Care Center During the COVID-19 Pandemic: Prospective Observational Study. J Med Internet Res 2020 Sep 21;22(9):e20874 [FREE Full text] [doi: 10.2196/20874] [Medline: 32896833]
7. Atherly A, Van Den Broek-Altenburg E, Hart V, Gleason K, Carney J. Consumer Reported Care Deferrals Due to the COVID-19 Pandemic, and the Role and Potential of Telemedicine: Cross-Sectional Analysis. JMIR Public Health Surveill 2020 Sep 14;6(3):e21607 [FREE Full text] [doi: 10.2196/21607] [Medline: 32833661]

XSL•FO
RenderX

8. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446. [doi: 10.1016/S2214-109X(15)00042-X]

9. Chan AT, Drew DA, Nguyen LH, Joshi AD, Ma W, Guo C, COPE Consortium. The COronavirus Pandemic Epidemiology (COPE) Consortium: A Call to Action. Cancer Epidemiol Biomarkers Prev 2020 Jul 05;29(7):1283-1289 [FREE Full text] [doi: 10.1158/1055-9965.EPI-20-0606] [Medline: 32371551]

10. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]

11. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8 [FREE Full text] [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]

12. Wissel B, Van Camp PJ, Kouril M, Weis C, Glauser TA, White PS, et al. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. J Am Med Inform Assoc 2020 Jul 01;27(7):1121-1125 [FREE Full text] [doi: 10.1093/jamia/ocaa071] [Medline: 32333753]

13. Cheng W, Hao C. Case-Initiated COVID-19 Contact Tracing Using Anonymous Notifications. JMIR Mhealth Uhealth 2020 Jun 22;8(6):e20369 [FREE Full text] [doi: 10.2196/20369] [Medline: 32501802]

14. Adorni F, Prinelli F, Bianchi F, Giacomelli A, Pagani G, Bernacchia D, et al. Self-Reported Symptoms of SARS-CoV-2 Infection in a Nonhospitalized Population in Italy: Cross-Sectional Study of the EPICOVID19 Web-Based Survey. JMIR Public Health Surveill 2020 Sep 18;6(3):e21866 [FREE Full text] [doi: 10.2196/21866] [Medline: 32650305]

15. Haodf.com. Webpage in Chinese. URL: https://www.haodf.com/ [accessed 2020-11-16]

16. Bao H, Cao B, Xiong Y, Tang W. Digital Media's Role in the COVID-19 Pandemic. JMIR mHealth uHealth 2020 Sep 18;8(9):e20156 [FREE Full text] [doi: 10.2196/20156] [Medline: 32530817]

17. Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 7). National Health Commission of the People's Republic of China. 2020 Mar 03. URL: https://www.chinadaily.com.cn/pdf/2020/1.Clinical.Protocols.for.the. Diagnosis.and.Treatment.of.COVID-19.V7.pdf [accessed 2020-11-16]

18. Guidelines for Investigation and Management of Close Contacts of COVID-19 Cases Training Kit from Chinese Center for Disease Control and Prevention. China Center for Disease Control and Prevention. 2020 Feb. URL: http://www. chinacdc.cn/en/COVID19/202003/P020200323390122945167.pdf [accessed 2020-11-16]

19. https://www.scmp.com/economy/china-economy/article/3064625/coronavirus-first-chinese-province-declare-top-level. South China Morning Post. 2020 Apr 08. URL: https://www.scmp.com/economy/china-economy/article/3064625/ coronavirus-first-chinese-province-declare-top-level [accessed 2020-11-16]

20. Dzieciatkowski T, Szarpak L, Filipiak KJ, Jaguszewski M, Ladny JR, Smereka J. COVID-19 challenge for modern medicine. Cardiol J 2020 May 18;27(2):175-183. [doi: 10.5603/cj.a2020.0055]

21. Feretzakis G, Kalles D, Verykios V. Hiding Decision Tree Rules in Medical Data: A Case Study. Stud Health Technol Inform 2019 Jul 04;262:368-371. [doi: 10.3233/SHTI190095] [Medline: 31349344]

22. Shen J, Duan H, Zhang B, Wang J, Ji JS, Wang J, et al. Prevention and control of COVID-19 in public transportation: Experience from China. Environ Pollut 2020 Nov;266(Pt 2):115291. [doi: 10.1016/j.envpol.2020.115291] [Medline: 32829124]

23. Ramtekkar U, Bridge JA, Thomas G, Butter E, Reese J, Logan E, et al. Pediatric Telebehavioral Health: A Transformational Shift in Care Delivery in the Era of COVID-19. JMIR Ment Health 2020 Sep 18;7(9):e20157 [FREE Full text] [doi: 10.2196/20157] [Medline: 32525485]

24. Sacco G, Lléonart S, Simon R, Noublanche F, Annweiler C, TOVID Study Group. Communication Technology Preferences of Hospitalized and Institutionalized Frail Older Adults During COVID-19 Confinement: Cross-Sectional Survey Study. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e21845 [FREE Full text] [doi: 10.2196/21845] [Medline: 32896832]

25. Mani VR, Kalabin A, Valdivieso SC, Murray-Ramcharan M, Donaldson B. New York Inner City Hospital COVID-19 Experience and Current Data: Retrospective Analysis at the Epicenter of the American Coronavirus Outbreak. J Med Internet Res 2020 Sep 18;22(9):e20548 [FREE Full text] [doi: 10.2196/20548] [Medline: 32540837]

Abbreviations

AI: artificial intelligence
GD2H: Guangdong Second Provincial General Hospital

XSL·FO
RenderX

Original Paper

Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach

Jia Xue[1,2], PhD; Junxiang Chen[3], PhD; Ran Hu[1], MSW, MA; Chen Chen[4], PhD; Chengda Zheng[2], BCom; Yue Su[5,6]; Tingshao Zhu[5], PhD

[1]Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

[2]Faculty of Information, University of Toronto, Toronto, ON, Canada

[3]School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[4]Middleware System Research Group, University of Toronto, Toronto, ON, Canada

[5]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[6]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Corresponding Author:
Tingshao Zhu, PhD
CAS Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
16 Lincui Road, Chaoyang District
Beijing, 100101
China
Phone: 86 0106485166
Email: tszhu@psych.ac.cn

Abstract

Background: It is important to measure the public response to the COVID-19 pandemic. Twitter is an important data source for infodemiology studies involving public response monitoring.

Objective: The objective of this study is to examine COVID-19–related discussions, concerns, and sentiments using tweets posted by Twitter users.

Methods: We analyzed 4 million Twitter messages related to the COVID-19 pandemic using a list of 20 hashtags (eg, "coronavirus," "COVID-19," "quarantine") from March 7 to April 21, 2020. We used a machine learning approach, Latent Dirichlet Allocation (LDA), to identify popular unigrams and bigrams, salient topics and themes, and sentiments in the collected tweets.

Results: Popular unigrams included "virus," "lockdown," and "quarantine." Popular bigrams included "COVID-19," "stay home," "corona virus," "social distancing," and "new cases." We identified 13 discussion topics and categorized them into 5 different themes: (1) public health measures to slow the spread of COVID-19, (2) social stigma associated with COVID-19, (3) COVID-19 news, cases, and deaths, (4) COVID-19 in the United States, and (5) COVID-19 in the rest of the world. Across all identified topics, the dominant sentiments for the spread of COVID-19 were anticipation that measures can be taken, followed by mixed feelings of trust, anger, and fear related to different topics. The public tweets revealed a significant feeling of fear when people discussed new COVID-19 cases and deaths compared to other topics.

Conclusions: This study showed that Twitter data and machine learning approaches can be leveraged for an infodemiology study, enabling research into evolving public discussions and sentiments during the COVID-19 pandemic. As the situation rapidly evolves, several topics are consistently dominant on Twitter, such as confirmed cases and death rates, preventive measures, health authorities and government policies, COVID-19 stigma, and negative psychological reactions (eg, fear). Real-time monitoring and assessment of Twitter discussions and concerns could provide useful data for public health emergency responses and planning. Pandemic-related fear, stigma, and mental health concerns are already evident and may continue to influence public trust when a second wave of COVID-19 occurs or there is a new surge of the current pandemic.

XSL•FO
RenderX

KEYWORDS

Introduction

Thirty million cases of COVID-19 have been confirmed across 110 countries as of mid-September 2020, and the death toll has reached close to 947,000 [1]. The widespread use of social media, such as Twitter, accelerates the process of exchanging information and expressing opinions about public events and health crises [2-5]. COVID-19 has been one of the trending topics on Twitter since January 2020 and has continued to be discussed to date. Since quarantine measures have been implemented across most countries (eg, the shelter-in-place order in the United States), people have been increasingly relying on different social media platforms to receive news and express opinions. Twitter data are valuable for revealing public discussions and sentiments related to various topics, as well as real-time news updates during global pandemics, such as H1N1 and Ebola [6-9]. Chew and Eysenbach's study [6] showed that Twitter could be used for real-time "infodemiology" studies, providing a source of opinions for health authorities to respond to public concerns. During the COVID-19 pandemic, many government officials worldwide have used Twitter as one of their main communication channels to regularly share policy updates and news related to COVID-19 to the general public [10].

Since the COVID-19 outbreak, a growing number of studies have collected Twitter data to understand the public responses to and discussions around COVID-19 [11-18]. For instance, Abd-Alrazaq and colleagues [11] adopted topic modeling and sentiment analysis to determine the main discussion themes and sentiments around COVID-19, using tweets collected between February 2 and March 15, 2020. Budhwani and Sun [14] compared Twitter discussions before and after March 16, 2020, when President Trump tweeted about the "Chinese virus," and found a significantly increased use of the phrase "Chinese virus" in people's tweets across many US states afterward. Mackey and colleagues [16] analyzed about 3465 tweets collected between March 2 and 20, 2020, using a topic model to explore users' self-reported experiences with COVID-19 and related symptoms. Ahmed and colleagues [12] conducted social network analysis and content analysis of collected tweets between March

27 and April 4, 2020, to understand what may have driven the misinformation that linked 5G towers in the United Kingdom to the COVID-19 pandemic. As conversations on Twitter continue to take place and evolve, it is worth continuing to use tweets as a source of data to track and understand the salient topics discussed on Twitter in response to the COVID-19 pandemic and track their changes across time.

To expand the literature on public reactions to the COVID-19 pandemic, this study aims to examine the public discourse and emotions related to the COVID-19 pandemic by analyzing more than 4 million tweets collected between March 7 and April 21, 2020.

Methods

Research Design

We used a purposive sampling approach to collect COVID-19–related tweets published between March 7 and April 21, 2020. Our Twitter data mining approach followed the pipeline displayed in Figure 1. Data preparation included the following three steps: (1) sampling, (2) data collection, and (3) preprocessing the raw data. The data analysis stage included unsupervised machine learning, sentiment analysis, and thematic qualitative analysis. The unit of analysis was each message-level tweet. Unsupervised learning is one approach in machine learning; it is used to examine data for patterns, and derives a probabilistic clustering based on text data. We chose unsupervised learning because it is commonly used when existing studies have few observations of or insights into unstructured text data [19]. Since a qualitative approach would be challenging when analyzing large-scale Twitter data, unsupervised learning allows us to conduct exploratory analyses of large text data for social science research. In this study, we first employed an unsupervised machine learning approach to identify salient latent topics. We used a thematic analysis approach to develop themes further, allowing a deeper dive into the data, such as through manual coding and inductively developing themes based on the latent topics generated by machine learning algorithms.

Figure 1. Twitter data mining pipeline.
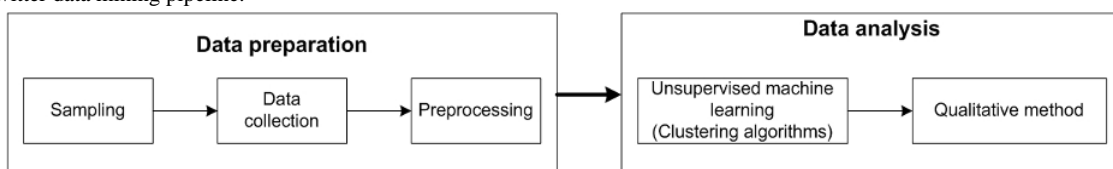
Sampling and Data Collection

We used a list of COVID-19–related hashtags as search terms to fetch tweets (eg, #coronavirus, #2019nCoV, #COVID19, #coronaoutbreak, and #quarantine; Multimedia Appendix 1). Twitter's open application programming interface (API) allowed us to collect updated Twitter messages set to open by default.

From March 7 to April 21, 2020, we collected 35,204,604 tweets during this period (Figure 2). After removing non-English tweets, 23,817,948 tweets remained. After removing duplicates and retweets (ie, tweets that only repost the original message without adding any more words), we had 4,196,020 tweets in our final data set. We collected and downloaded the following features for each tweet: (1) the full text, (2) the numbers of

favorites, followers, and followings, (3) users' geolocation, and (4) users' description/self-created profile.

Figure 2. Tweet preprocessing chart.
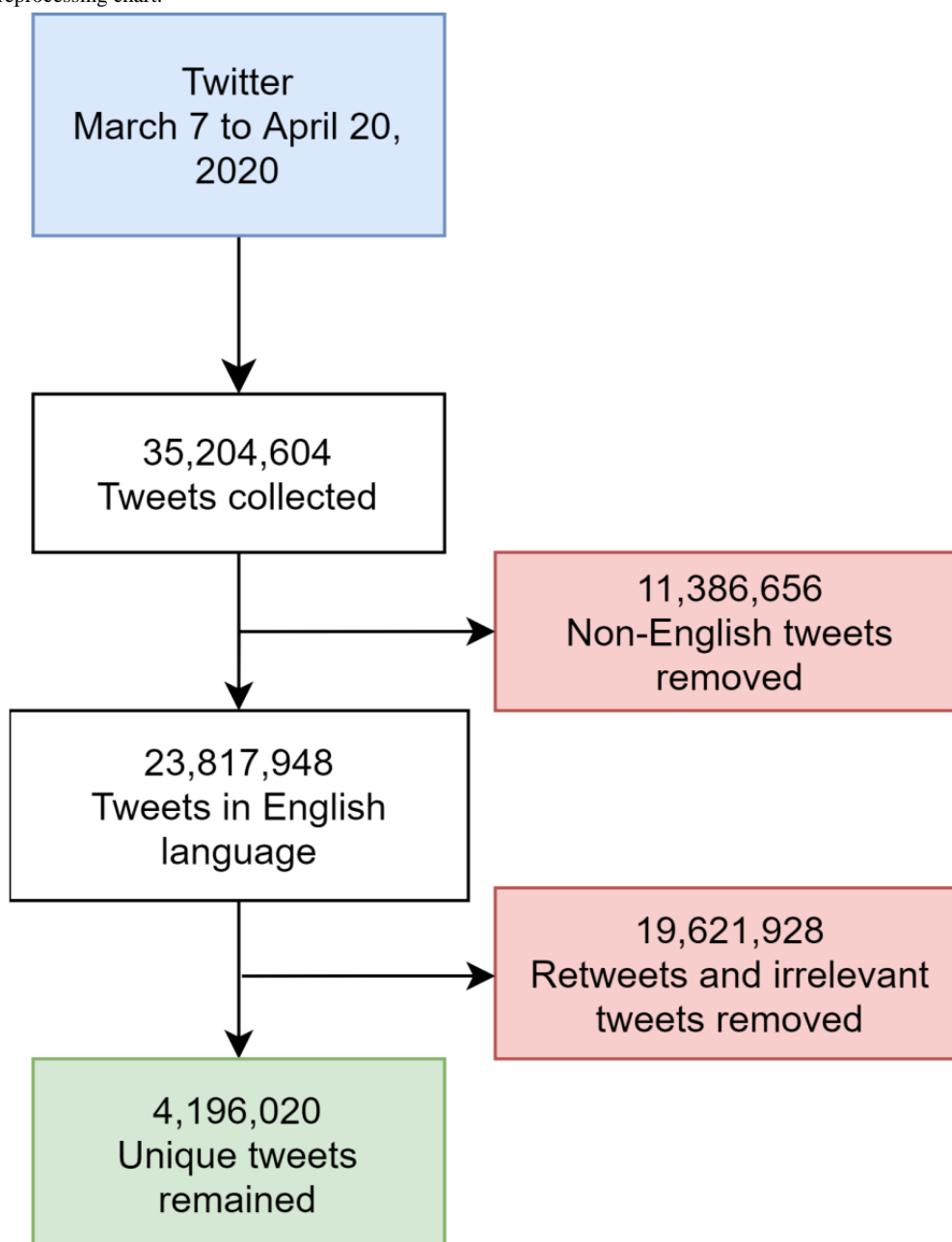
Preprocessing the Raw Data

We used Python to clean the raw data (Figure 1). The process was as follows [18]:

1. We removed the hashtag symbol, @users, and URLs from the tweets in the data set.
2. We removed non-English characters (non-ASCII characters) because this study focused on tweets in English.
3. We removed special characters, punctuation, and stop-words [19] from the data set as they do not contribute to the semantic meanings of messages.

Data Analysis

Unsupervised Machine Learning

Latent Dirichlet Allocation (LDA) [20] is a widely used unsupervised machine learning approach that allows researchers to analyze unstructured text data (eg, Twitter messages). Based on the data itself, the algorithm produces frequently mentioned pairs of words, the pairs of words that co-occur together, and latent topics and their distributions over topics in the document [21]. Existing studies have indicated the feasibility of using LDA to identify the patterns and themes of tweets related to COVID-19 [11,22].

Qualitative Analysis

To triangulate and contextualize findings from the LDA model, we employed a qualitative approach to develop themes further.

Specifically, we used Braun and Clarke's [23] six steps of thematic analysis: (1) getting familiar with the keyword data, (2) generating initial codes, (3) searching for themes, (4) reviewing potential themes, (5) defining themes, and (6) reporting. In addition to following the six-phase approach, our process was iterative and reflective by moving backward and forward through the six phases [24]. The thematic approach relied on human interpretation, a process that can be significantly influenced by personal understanding of the topics and a variety of biases. Two team members who have experience analyzing Twitter data documented their thoughts about potential codes in NVivo independently. Two other team members then reviewed the initial codes and considered whether they reflected the identified topics. For example, two team members collapsed several similar codes into one theme to ensure the topics corresponded meaningfully under one theme. The next stage was naming the themes to ensure the themes fitted into the overall meanings of the identified salient topics. We finalized themes corresponding to each of the 13 topics.

Sentiment Analysis

We used sentiment analysis, a natural language processing (NLP) approach, to classify the main sentiments of a given twitter message, such as fear and joy [25]. In this study, we used the NRC Emotion Lexicon, which consists of 8 primary emotions: anger, anticipation, fear, surprise, sadness, joy, disgust, and trust [26]. We followed 4 steps to calculate the emotion index for each Twitter message: (1) removed articles and pronouns (eg, "and," "the," "to"), (2) applied a stemmer by removing the predefined list of prefixes and suffixes (eg, "running" becomes "run" after stemming) [27], and (3) calculated the emotion index (if a sentence had multiple emotions, we only kept the emotion with the highest matching count), and (4) calculated the scores for each 8-emotion type. We discussed these 4 steps in detail in a previous study [18].

Results

Descriptive Results

In total, after preprocessing all raw data, our final data set included 4,196,020 tweets. We identified the most popular tweeted bigrams (pairs of words) related to COVID-19. Bigrams captured "two concessive words regardless of the grammar structure and semantic meaning and may not be self-explanatory" [21]. Bigrams identified included the following: "covid 19," "stay home," "social distancing," "new cases," "don't know," "confirmed cases," "home order," "New York," "tested positive," "death toll," and "stay safe." Popular unigrams included "virus," "lockdown," "quarantine," "people," "new," "home," "like," "stay," "don't," and "cases." We presented the most popular unigrams and bigrams related to COVID-19 in Table 1 and visualized them using word clouds in Figures 3 and 4.

Table 1. Top 50 bigrams and unigrams and their distributions.

Top 50 bigrams	Percentage of data set	Top 50 unigrams	Percentage of data set
covid 19	0.29	virus	1.18
stay home	0.26	lockdown	0.98
corona virus	0.12	quarantine	0.94
social distancing	0.08	people	0.82
new cases	0.07	coronavirus	0.79
dont know	0.04	new	0.47
confirmed cases	0.04	home	0.45
home order	0.04	like	0.44
new york	0.04	im	0.41
tested positive	0.04	stay	0.41
death toll	0.04	dont	0.41
home orders	0.04	cases	0.37
quarantine got	0.03	time	0.36
stay safe	0.03	covid	0.35
spread virus	0.03	19	0.30
coronavirus cases	0.03	need	0.30
shelter place	0.03	day	0.29
coronavirus pandemic	0.03	trump	0.28
year old	0.03	china	0.28
public health	0.03	know	0.28
chinese virus	0.03	going	0.25
ill deliver	0.03	help	0.25
deliver copy	0.03	pandemic	0.24
health care	0.03	world	0.24
support usps	0.03	health	0.23
signing support	0.02	think	0.22
usps ill	0.02	deaths	0.21
wuhan virus	0.02	today	0.21
quarantine im	0.02	good	0.20
mental health	0.02	work	0.20
dont want	0.02	want	0.19
im going	0.02	corona	0.17
president trump	0.02	spread	0.17
united states	0.02	got	0.17
dont think	0.02	support	0.17
copy officials	0.02	government	0.17
feel like	0.02	right	0.15
looks like	0.02	way	0.15
positive cases	0.02	care	0.15
staying home	0.02	social	0.15
officials toodelivered	0.02	news	0.15
coronavirus outbreak	0.02	state	0.15

XSL•FO
RenderX

Top 50 bigrams	Percentage of data set	Top 50 unigrams	Percentage of data set
domestic violence	0.02	country	0.15
coronavirus lockdown	0.02	said	0.14
healthcare workers	0.02	ive	0.14
people died	0.02	days	0.14
quarantine day	0.02	testing	0.14
donald trump	0.02	stop	0.13
social media	0.02	says	0.13

Figure 3. The word cloud of the most popular unigram.

Figure 4. Word cloud of the most popular bigrams.

COVID-19–Related Topics

Our approach, LDA, produced frequently co-occurring pairs of words related to COVID-19 and organized these co-occurring words into different topics. LDA allowed us to manually define the number of topics (eg, 10 topics, 20 topics) that we would like to generate. Consistent with previous studies, we used the coherence model, Gensim (RARE Technologies Ltd) [28], to calculate the most appropriate number of topics based on the data itself. For this data set, the LDA indicated that having 13 topics would give a high coherence score and the smallest topic number (eg, while having 19 or 20 topics would give a higher coherence score, they involve more topics; Figure 5).

Figure 5. The number of topics based on the coherence model.

We further analyzed the document-term matrix and obtained the distributions of 13 topics. We presented the results of 13 salient topics and the most popular pairs of words (bigrams) within each topic in Table 2. For example, Topic 3 had the highest distribution (8.87%) among all 13 common latent topics.

The bigrams associated with Topic 3 included "tested positive," "coronavirus outbreak," "New York," "shelter place," and "mental health." These pairs of words frequently co-occurred together, and therefore the LDA model assigned them to the same topic.

Table 2. Identified salient topics, bigrams, and their distributions.

Topic	Bigrams within topics	Distribution (%)
1	covid 19, dont know, deadly virus, im gonna, spreading virus, 19 lockdown, herd immunity, 000 people, 19 pandemic, dont need, face masks, fox news, health workers, small businesses, home quarantine, like this, virus came, slow spread, test kits, total confirmed	8.51
2	spread virus, health care, staying home, white house, positive cases, people die, 14 days, coronavirus deaths, care workers, ive seen, need help, day lockdown, know virus, im getting, doctors nurses, quarantine period, virus world, stop virus, people getting, week quarantine	7.24
3	tested positive, coronavirus outbreak, wuhan virus, positive coronavirus, confirmed cases, new york, shelter place, mental health, china virus, feel like, new cases, gt gt, coronavirus covid, virus, weeks, people virus, people don't, bringing total, press conference, sars cov	8.87
4	dont think, virus spread, lockdown period, fake news, nursing homes, wuhan lab, best thing, months, lockdown amp, 21 3, id like, people know, real time, entire world, know im, know it, wake up, feel free, dont wanna, anthony fauci	6.56
5	u s, coronavirus cases, public health, save lives, novel coronavirus, long term, south korea, dont forget, bbc news, care homes, news coronavirus, million people, doesnt mean, family members, want know, coronavirus vaccine, going on, rest world, coronavirus, new jersey	7.36
6	at home, stay at, home order, thank you, look like, good news, test positive, people stay, fight virus, people protesting, face mask, good thing, young people, lock down, wearing masks, cases deaths, trump said, deaths reported, shut down, active cases	7.36
7	social distancing, day quarantine, healthcare workers, prime minister, world health, dont care, global pandemic, dont understand, health organization, dr fauci, let know, time lockdown, virus isn't, in place, anti lockdown, shelter in, people think, live updates, 2 months	7.81
8	coronavirus lockdown, coronavirus crisis, amid coronavirus, looks like, new coronavirus, task force, im sure, coronavirus patients, prevent spread, virus doesn't, dont let, long time, new York, high risk, coronavirus task, thank god, number deaths, dont like, virus outbreak, coronavirus cases	7.47
9	stay safe, chinese virus, self quarantine, need know, people going, new virus, common sense, safe stay, virus amp, b c, 2 2, family friends, we've got, got virus, stay away, testing kits, health amp, virus gone, april 20, knew virus	7.07
10	corona virus, new cases, death toll, im going, quarantine day, people died, spread coronavirus, cases coronavirus, people dying, quarantine im, total number, number cases, cases reported, april 2020, confirmed cases, coronavirus death, 24 hours, people need, stop spread	8.84
11	stay home, home orders, president trump, social media, home stay, loved ones, stay safe, death rate, working home, 31 000, social distance, 3100 000, protesting stay, breaking news, deaths, im sorry, 10 000, mortality rate	8.67
12	coronavirus pandemic, year old, united states, wash hands, people like, work home, god bless, lot people, wear mask, years ago, virus hoax, like virus, 23 days, grocery store, said virus, 21 million, watch video, 10 days, like amp, uk lockdown	7.06
13	right now, dont want, 3 weeks, tests positive, donald trump, weeks ago, weeks lockdown, virus spreading, coronavirus update, new zealand, 22 million, sounds like, total cases, lockdown 2, communist party, day day, chinese communist, cases 1, whats happening, 2 weeks	7.18

COVID-19–Related Themes

The thematic analysis enabled us to categorize these topics into different distinct themes. The team considered the identified topics, bigrams, and representative tweet samples in each topic and categorized them into different themes. To protect the privacy and anonymity of the Twitter users, we did not present any user-related information, such as users' Twitter handles or other identifying information. Therefore, sample tweets were excerpts drawn from original tweets in Table 3.

We organized 13 topics into 5 themes: "Public health measures to slow the spread of COVID-19" (eg, face masks, test kits, vaccine), "Social stigma associated with COVID-19" (eg, Chinese virus, Wuhan virus), "Coronavirus news cases and deaths" (eg, new cases, deaths), "COVID-19 in the United States" (eg, New York, protests, task force), and "Coronavirus cases in the rest of the world" (eg, UK, global issue). For example, the theme "public health measures to slow the spread of COVID-19" included the relevant topics of "facemasks," "quarantine," "test kits," "lockdown," "safety," "vaccine," and "shelter-in-place." In addition, "home quarantine" and "self-quarantine" were two of the most commonly co-occurred words under the topic quarantine.

XSL•FO
RenderX

Table 3. Themes based on topic classification, bigrams, and sample tweets.

Theme and topic	Bigrams	Sample tweets
Public health measures to slow the spread of COVID-19		
Face masks	face masks, wear masks	We protect us and our family by wearing masks every day.
Quarantine	home quarantine, self quarantine, quarantine period	@realDonaldTrump @JustineTrudeau They're all under mandatory 2 week quarantine, and they are essential workers…
Test kits	test kits, testing kits	Hydroxychloroquine, Testing Kits and USA: We urge the Modi govt to draw proper lessons from this latest instance of US
Lockdown	covid19 lockdown, lockdown period, weeks lockdown,	People are actually shocked the lockdown has been extended for 3 weeks when there are still people going out meeting
Safety	stay safe, safe stay, stay away	Be strong, stay safe #lockdown but not locked out http://t.co/Fvi-fiEbbs7
Vaccine	coronavirus vaccine	Lead scientist for NIH working on #coronavirus vaccine research
US shelter-in-place	Shelter place, shelter in	Did California's shelter-in-place order work? If you sue crap data without any reference to epidemiology, then yes
Social stigma associated with COVID-19		
Chinese Communist Party	communist party, Chinese communist, cases 1	The #Chinese Communist Party (#CCP) is spreading disinformation to cover up the origin of the #coronavirus
Discriminatory names	Wuhan virus, Chinese virus	That China is responsible for putting entire world @great risk. Heavily criticized their eating habits.
President Trump tweeting "Chinese virus"	president trump, social media, china virus	President Trump: They know where it came from. We all know where it came from, #chinesevirus
COVID-19 new cases and deaths		
New cases	new cases, total number, confirmed cases	RT @neeratanden: 4,591 people died in a day from the virus, the highest number anywhere ever that we know of.
Deaths	coronavirus death, death toll, people died	#Britain's death toll could be DOUBLE official tally as care homes
COVID-19 in the United States		
Mental health and COVID-19 in New York	new york, shelter place, mental health	New Yorkers on their apartment roofs during quarantine is a whole different vibe. This is gonna be in history books
Protests against the lockdown	anti lockdown, people protesting, protesting stay	I stand with the Healthcare workers!!! Bravo! Healthcare workers face off against anti-lockdown protesters in Colorado
Task force in the United States	task force	RT @Jim_Jordan: There are #coronavirus task forces doing great work. But there is one task force that's missing in action: the U.S. congress
COVID-19 pandemic in the United States	united states, white house, new jersey, 21 million, million people, dr fauci,	Stay-at-home orders continue in much of the United States
COVID-19 cases in the rest of the world		
United Kingdom	Herd immunity, UK lockdown, Prime Minister	The Prime Minister gave the game away early on when he openly said to Scrofulous and Willibooby that the government's plan was Herd Immunity the REAL people in charge must have been so furious with him he had to be sent to an isolation ward with the virus to shut him up!
Global issue	Entire world, south Korea, world health, global pandemic, new Zealand	Worldwide it is now 182,726." And "New Zealand Prime Minster Jacinda Ardern says the government will partially relax its lockdown in a week, as a decline in …

Sentiment Analysis

We presented the results of the sentiment analysis for each of the 13 latent topics in Figure 6 and Table 4. Figure 6 presented 8 emotions of trust, anticipation, joy, surprise, anger, fear, disgust, and sadness. Results showed that across all 13 topics, anticipation (dark blue line) dominated 12 topics, followed by fear (orange line), trust (grey line), and anger (yellow line).

We also ran a one-tailed z test to examine if each of the 8 emotions is statistically significantly different across topics. A P value <.01 was set as the threshold for significance. For example, about 23.8% of tweets in Topic 5 revealed feelings of

anticipation that "necessary steps and precautions will be taken" [18,29]. Statistical significance indicated that it was very likely ($P<.001$) that the anticipation emotion is more prevalently expressed in Topic 5 (23.8%) than all other topics. The emotion fear (of the impacts of the virus) was found in 18.8% of the tweets in Topic 10, which was statistically different from the fear expressed in other topics.

Figure 6. Sentiment analysis for each of the 13 latent topics.
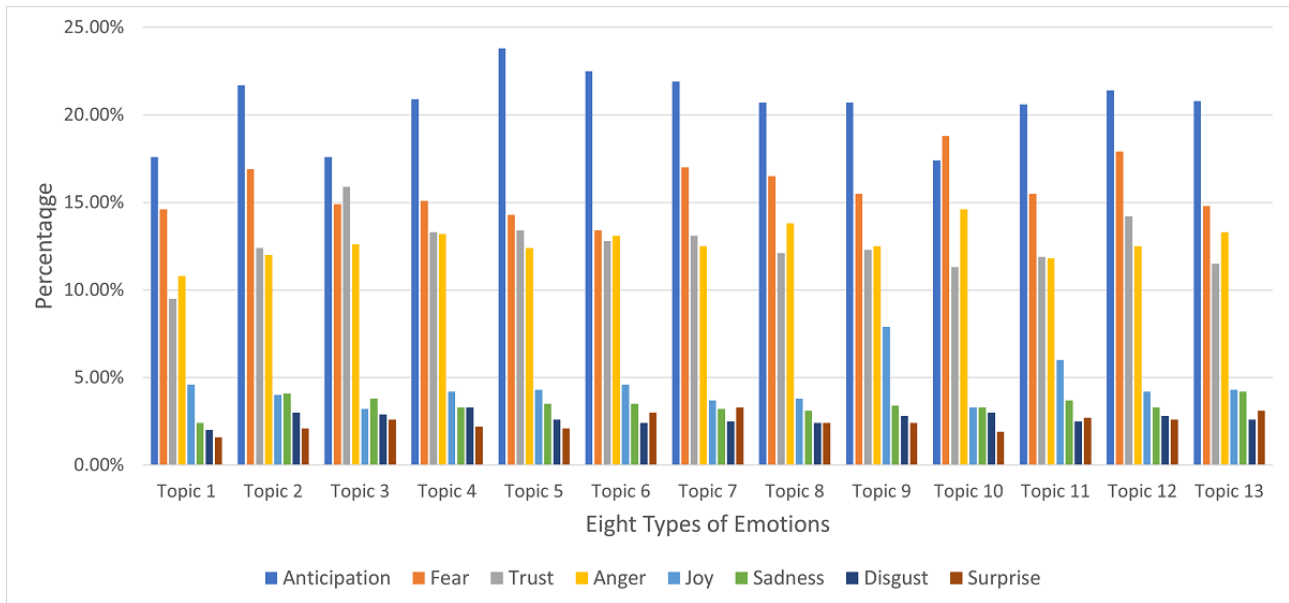
Table 4. The percentage of 8 emotions across 13 topics[a].

Topic	Anger, %	Anticipation, %	Disgust, %	Fear , %	Joy, %	Sadness, %	Surprise, %	Trust, %
1	10.80	17.60	2.00	14.60	4.60[b]	2.40	1.60	9.50
2	12.00	21.70[b]	3.00[b]	16.90[b]	4.00	4.10[b]	2.10	12.40[b]
3	12.60[b]	17.60	2.90[b]	14.90	3.20	3.80[b]	2.60[b]	15.90[b]
4	13.20[b]	20.90[b]	3.30[b]	15.10	4.20	3.30[b]	2.20[b]	13.30[b]
5	12.40[b]	23.80[b]	2.60[b]	14.30	4.30	3.50[b]	2.10	13.40[b]
6	13.10[b]	22.50[b]	2.40	13.40	4.60[b]	3.50[b]	3.00[b]	12.80[b]
7	12.50[b]	21.90[b]	2.50[b]	17.00[b]	3.70	3.20[b]	3.30[b]	13.10[b]
8	13.80[b]	20.70[b]	2.40	16.50[b]	3.80	3.10[b]	2.40[b]	12.10[b]
9	12.50[b]	20.70[b]	2.80[b]	15.50	7.90[b]	3.40[b]	2.40[b]	12.30[b]
10	14.60[b]	17.40	3.00[b]	18.80[b]	3.30	3.30[b]	1.90	11.30
11	11.80	20.60[b]	2.50[b]	15.50[b]	6.00[b]	3.70[b]	2.70[b]	11.90[b]
12	12.50[b]	21.40[b]	2.80[b]	17.90[b]	4.20	3.30[b]	2.60[b]	14.20[b]
13	13.30[b]	20.80[b]	2.60[b]	14.80	4.30	4.20[b]	3.10[b]	11.50[b]

[a]The sum of the percentages for each topic is not equal to 100%. The remainder is made up of neutral or other emotions.

[b]$P<.001$ from z test.

Discussion

Principal Results

In this study, we addressed public discussions and emotions using COVID-19–related messages on Twitter. Twitter users discussed 5 main themes related to COVID-19 between March 7 and April 21, 2020. Topic modeling of the tweets was useful for providing insights about COVID-19 topics and concerns. Results showed several essential points. First, the public uses a variety of terms when referring to COVID-19, including virus, COVID-19, coronavirus, and corona virus. Second, COVID-19 has been referred to as the "China virus," which can create stigma and harm efforts to address the COVID-19 outbreak

[14]. Third, discussions about the pandemic in New York were salient, and its associated public sentiment was anger. Fourth, public discussions about the Chinese Communist Party (CCP) and the spread of the virus emerged as a new topic that was not identified in previous studies [18], suggesting the connection between COVID-19 and politics is increasingly circulating on Twitter as the situation evolves. Fifth, public sentiments on the spread of COVID-19 reveal anticipation for the potential measures that can be taken, followed by mixed feelings of trust, anger, and fear. Results suggest that the public is not surprised by the rapid spread of COVID-19. Sixth, people have a significant feeling of fear when they discuss the COVID-19 crisis and deaths. Lastly, trust is no longer a prominent emotion when Twitter users discuss COVID-19, which is different from the findings of an earlier study [18].

Comparison With Prior Work

Our findings are consistent with previous studies using social media data to assess the public health responses and sentiments related to COVID-19, and suggest that public attention has been focusing on the following topics since January 2020: (1) the confirmed cases and death rates [11,18,30], (2) preventive measures [11,18,31], (3) health authorities and government policies [10,18], (4) an outbreak in New York [18], (5) COVID-19 stigma (eg, referring to COVID-19 as the "Chinese virus") [11,14], and (6) negative psychological reactions (eg, fear) or mental health consequences [11,31-33].

Compared with a study examining public discussions and concerns related to COVID-19 using Twitter data from January 20 to March 7, 2020, we found that several salient topics are no longer popular: (1) an outbreak in South Korea, (2) the *Diamond Princess* cruise ship, (3) the economic impact [11,32], and (4) supply chains [18]. Given current preventive measures, washing hands is no longer a prevalent topic; instead, quarantine has become dominant.

In addition, our study identified new discussion topics about COVID-19 occurring between March 7 to April 21: (1) the need for a vaccine to stop the spread, (2) quarantine and shelter-in-place orders, (3) protests against the lockdown, and (4) the COVID-19 pandemic in the United States. The new salient topics suggest that Twitter users (tweeting in English) are focusing their attention on COVID-19 in the United States (eg, New York, protests, task force, millions of confirmed cases) rather than global news (eg, South Korea, *Diamond Princess* cruise ship, Dr Li Wenliang in China).

Limitations

First, we only sampled 20 hashtags as the key search terms to collect Twitter data (Multimedia Appendix 1). New hashtags keep coming up as the situation evolves. For example, a hashtag may become widely used after a related topic becomes more popular, such as the official name for the virus (COVID-19).

Second, Twitter users are not representative of the whole global population, and topics of tweets only indicate online users' opinions about and reactions to COVID-19. However, the Twitter data set is still a valuable resource, allowing us to examine real-time Twitter users' responses and online activities related to COVID-19. Third, non-English tweets were removed from our analyses, and hence the results are limited to users who posted in English only. Future COVID-19 studies should include other languages, such as Italian, French, German, and Spanish.

Future Research

Future research could further explore public trust and confidence in existing measures and policies, which are essential. Compared to prior work, our study showed that Twitter users had a feeling of joy when talking about herd immunity. Sentiments of fear and anticipation related to the topics of quarantine and shelter-in-place. Future studies could evaluate how government officials (eg, President Trump) and international organizations (eg, World Health Organization) deliver and convey messages to the public, and the subsequent impact on public opinions and sentiments. Anti-Chinese/Asian sentiments spread on social media, and it would be worth assessing how people use these platforms to resist and challenge COVID-19 stigma. Misinformation during the COVID-19 pandemic was not a prominent theme in this study. An existing study showed that 25% (n=153) of sampled tweets contained misinformation [34]. The term COVID-19 has lower rates of misinformation associated with it than that associated with #2019_ncov and Corona. Future research should investigate misinformation and how it expands on social media. Finally, trust is no longer prominent when people tweet about confirmed cases and deaths. Instead, fear has replaced trust to be the dominant emotion. Future research should examine the changes in trust over time.

Conclusions

Twitter data and machine learning approaches can be leveraged for infodemiology studies by studying evolving public discussions and sentiments during the COVID-19 pandemic. Our findings facilitate an understanding of public discussions and concerns about the COVID-19 pandemic among Twitter users between March 7 and April 21, 2020. Several topics were consistently dominant on Twitter, such as "the confirmed cases and death rates," "preventive measures," "health authorities and government policies," "stigma," and "negative psychological reactions" (eg, fear). As the situation rapidly evolves, new salient topics emerge accordingly. Fear arises in messages of new cases or death reports [18]. Real-time monitoring and assessment of Twitter users' concerns can be promising for informing public health emergency responses and planning. Hearing and reacting to real concerns from the public can enhance trust between the health care system and the public and enable better preparation for a future public health emergency.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 13 KB - jmir_v22i11e20550_app1.docx]

References

1. Center for Systems Science and Engineering (CSSE). COVID-19 Dashboard by CSSE at Johns Hopkins University (JHU). URL: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-06-16]

2. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]

3. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]

4. Xue J, Chen J, Chen C, Hu R, Zhu T. The Hidden Pandemic of Family Violence During COVID-19: Unsupervised Learning of Tweets. J Med Internet Res 2020 Nov 06;22(11):e24361-e24353 [FREE Full text] [doi: 10.2196/24361] [Medline: 33108315]

5. Cheong M, Lee VCS. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Inf Syst Front 2010 Sep 29;13(1):45-59. [doi: 10.1007/s10796-010-9273-x]

6. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]

7. Jones JH, Salathé M. Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One 2009 Dec 03;4(12):e8032 [FREE Full text] [doi: 10.1371/journal.pone.0008032] [Medline: 19997505]

8. Kim Y, Kim JH. Using photos for public health communication: A computational analysis of the Centers for Disease Control and Prevention Instagram photos and public responses. Health Informatics J 2020 Sep;26(3):2159-2180 [FREE Full text] [doi: 10.1177/1460458219896673] [Medline: 31969051]

9. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

10. Rufai S, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. J Public Health (Oxf) 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: 10.1093/pubmed/fdaa049] [Medline: 32309854]

11. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

12. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]

13. Alvarez-Risco A, Mejia C, Delgado-Zegarra J, Del-Aguila-Arcentales S, Arce-Esquivel A, Valladares-Garrido M, et al. The Peru Approach against the COVID-19 Infodemic: Insights and Strategies. Am J Trop Med Hyg 2020 Aug;103(2):583-586 [FREE Full text] [doi: 10.4269/ajtmh.20-0536] [Medline: 32500853]

14. Budhwani H, Sun R. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the. J Med Internet Res 2020 May 06;22(5):e19301. [doi: 10.2196/19301] [Medline: 32343669]

15. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]

16. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]

17. Xiang X, Lu X, Halavanau A, Xue J, Sun Y, Lai P, et al. Modern Senicide in the Face of a Pandemic: An Examination of Public Discourse and Sentiment about Older Adults and COVID-19 Using Machine Learning. J Gerontol B Psychol Sci Soc Sci 2020 Aug 12:A [FREE Full text] [doi: 10.1093/geronb/gbaa128] [Medline: 32785620]

18. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]

19. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. New York, NY, USA: Springer; 2013.

20. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research March 2003;3:993-1022 [FREE Full text]

21. Xue J, Chen J, Gelles R. Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter. Violence and Gender 2019 Jun;6(2):105-114. [doi: 10.1089/vio.2017.0066]

22. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]

XSL•FO
RenderX

23.   Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

24.   Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis. International Journal of Qualitative Methods 2017 Oct 02;16(1):1. [doi: 10.1177/1609406917733847]

25.   Beigi G, Hu X, Maciejewski R, Liu H. An overview of sentiment analysis in social mediaits applications in disaster relief. In: Pedrycz W, Chen SM, editors. Studies in Computational Intelligence. New York City, NY, USA: Springer Verlag; 2016:313-340.

26.   Mohammad S, Turney P. NRC emotion lexicon. National Research Council Canada. 2013. URL: https://nrc-publications. canada.ca/eng/view/object/?id=0b6a5b58-a656-49d3-ab3e-252050a7a88c [accessed 2020-11-12]

27.   Manning C, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press; 2008.

28.   Michael R, Andreas B, Alexander H. Exploring the space of topic coherence measures. 2015 Presented at: Eighth ACM International Conference on Web Search and Data Mining; 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]

29.   Kaila RP, Prasad AVK. Informational flow on Twitter - corona virus outbreak - topic modelling approach. International Journal of Advanced Research in Engineering and Technology 2020;11(3):128-134 [FREE Full text]

30.   Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public Priorities and Concerns Regarding COVID-19 in an Online Discussion Forum: Longitudinal Topic Modeling. J Gen Intern Med 2020 Jul 12;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]

31.   Nobles J, Martin F, Dawson S, Moran P, Savovic J. The potential impact of COVID-19 on mental health outcomes and the implications for service solutions. 2020 Apr 15. URL: https://arc-w.nihr.ac.uk/research-and-implementation/covid-19-response/reports/potential-impact-of-covid-19-on-mental-health-outcomes-and-the-implications-for-service-solutions/ [accessed 2020-11-12]

32.   Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]

33.   Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):4552 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]

34.   Medford R, Saleh S, Sumarsono A, Perl T, Lehmann C. An. Open Forum Infect Dis 2020 Jul;7(7):ofaa258 [FREE Full text] [doi: 10.1093/ofid/ofaa258] [Medline: 33117854]

Abbreviations

API: Application Programming Interface
CCP: Chinese Communist Party
LDA: Latent Dirichlet Allocation
NLP: Natural Language Processing

Original Paper

Chinese Residents' Perceptions of COVID-19 During the Pandemic: Online Cross-sectional Survey Study

Tingting Cui[1,2*], MA; Guoping Yang[3*], MM; Lili Ji[3], MM; Lin Zhu[3], MM; Shiqi Zhen[3], PhD; Naiyang Shi[1,2], MM; Yan Xu[3], MM; Hui Jin[1,2], PhD

[1]Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, China

[2]Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing, China

[3]Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Hui Jin, PhD
Department of Epidemiology and Health Statistics
School of Public Health
Southeast University
No. 87 Dingjiaqiao
Nanjing,
China
Phone: 86 025 8327 2572
Fax: 86 825 8327 2561
Email: jinhuihld@seu.edu.cn

Abstract

Background: COVID-19 has posed a global threat due to substantial morbidity and mortality, and health education strategies need to be adjusted accordingly to prevent a possible epidemic rebound.

Objective: This study aimed to evaluate the perceptions of COVID-19 among individuals coming to, returning to, or living in Jiangsu Province, China, and determine the impact of the pandemic on the perceptions of the public.

Methods: In this study, an online questionnaire was distributed to participants between February 15 and April 21, 2020. The questionnaire comprised items on personal information (eg, sex, age, educational level, and occupation); protection knowledge, skills, and behaviors related to COVID-19; access to COVID-19–related information; and current information needs. Factors influencing the knowledge score, skill score, behavior score, and total score for COVID-19 were evaluated using univariate and multivariate analyses. The time-varying reproduction number (R_t) and its 95% credible interval were calculated and compared with the daily participation number and protection scores.

Results: In total, 52,066 participants were included in the study; their average knowledge score, skill score, behavior score, and total score were 25.58 (SD 4.22), 24.05 (SD 4.02), 31.51 (SD 2.84), and 90.02 (SD 8.87), respectively, and 65.91% (34,315/52,066) had a total protection score above 90 points. For the knowledge and skill sections, correct rates of answers to questions on medical observation days, infectiousness of asymptomatic individuals, cough or sneeze treatment, and precautions were higher than 95%, while those of questions on initial symptoms (32,286/52,066, 62.01%), transmission routes (37,134/52,066, 71.32%), selection of disinfection products (37,390/52,066, 71.81%), and measures of home quarantine (40,037/52,066, 76.90%) were relatively low. For the actual behavior section, 97.93% (50,989/52,066) of participants could wear masks properly when going out. However, 19.76% (10,290/52,066) could not disinfect their homes each week, and 18.42% (9589/52,066) could not distinguish differences in initial symptoms between the common cold and COVID-19. The regression analyses showed that the knowledge score, skill score, behavior score, and total score were influenced by sex, age, educational level, occupation, and place of residence at different degrees ($P<.001$). The government, television shows, and news outlets were the main sources of protection knowledge, and the information released by the government and authoritative medical experts was considered the most reliable. The current information needs included the latest epidemic developments, disease treatment progress, and daily protection knowledge. The R_t in the Jiangsu Province and mainland China dropped below 1, while the global R_t remained at around 1. The maximal information coefficients ranged from 0.76 to 1.00, which indicated that the public's perceptions were significantly associated with the epidemic.

XSL·FO
RenderX

Conclusions: A high proportion of the participants had sufficient COVID-19 protection knowledge and skills and were able to avoid risky behaviors. Thus, it is necessary to apply different health education measures tailored to work and study resumption for specific populations to improve their self-protection and, ultimately, to prevent a possible rebound of COVID-19.

Introduction

COVID-19, caused by the novel coronavirus SARS-CoV-2, was first reported in late December 2019 in Wuhan City, China. As a relative of the two conditions caused by previous coronaviruses, namely, the severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), COVID-19 has posed a great global threat due to its substantial morbidity and mortality [1,2], and was declared by the World Health Organization (WHO) to be a public health emergency of international concern on January 30, 2020 [3]. Preliminary research has shown that the severity of COVID-19 was lower than that of SARS and MERS; however, it may be more infectious [2,4,5]. According to the WHO's COVID-19 situation reports [6,7], as of February 15, 2020, a total of 51,857 cases and 1669 deaths were confirmed in only 26 countries, areas, or territories; as of April 21, 2020, the total number of confirmed cases had increased to 2,471,136, with 169,006 deaths, in over 200 countries, areas, or territories. At present, a variety of candidate vaccines have been developed or are undergoing clinical trials to control the pandemic [8,9].

It is essential to pay attention to the public's knowledge level, attitudes, and perceptions in order to customize the preventive and control measures applied by governments and health authorities during rapidly spreading infectious disease outbreaks [10,11]. During the early period of the COVID-19 outbreak, most participants in the investigations [11-14] conducted in the United States, the United Kingdom, China, and Ethiopia had certain knowledge on COVID-19, such as the main transmission routes and common symptoms, with optimistic attitudes and appropriate practices. However, misconceptions on how to prevent an infection and recommended care-seeking behaviors still existed [11]. Similarly, the knowledge and practices required to combat COVID-19 among high-risk populations are insufficient [14]. Disease perception then plays a relevant role in individuals' psychological adjustment. The Brief Illness Perception Questionnaire (BIP-Q5) was used to measure the psychological impact during the COVID-19 outbreak in a sample of Spanish adults [15], which showed adequate psychometric properties.

The COVID-19 epidemic in China has been essentially controlled, and the resumption of work and study is proceeding in an orderly manner. Jiangsu Province, located in the eastern part of China, has a large labor import and rich educational resources. Under this scenario, the cross-regional movement of individuals will lead to an increased risk of epidemic imports as well as cluster transmissions of COVID-19. An investigation on the status of protection against COVID-19 among individuals coming to, returning to, or living in Jiangsu Province will help provide information on the current mastery level of knowledge, skills, and protection behaviors; popularize prevention and control knowledge; and tailor health education strategies in a timely manner to ultimately prevent a possible epidemic rebound.

Methods

Study Design and Participants

In this study, a cross-sectional online survey was conducted on the public platform created by Jiangsu Provincial People's Government and managed by the Jiangsu Provincial Center for Disease Control and Prevention (CDC). The platform included Jiangsu health codes, through which every citizen had the obligation to fill in health information during the COVID-19 outbreak; otherwise, they were not permitted to enter/exit public places and their workplaces [16]. The platform included approximately 30,000,000 participants. The questionnaire used in this study has been embedded on the platform since February 15, 2020, participation was anonymous and voluntary, and participants had one chance to fill in their information. This study was approved by the Ethics Committee of the Jiangsu Provincial CDC.

Data Collection

Data were collected using an online questionnaire through WeChat (Multimedia Appendix 1). The questionnaire was created according to the national guideline for the diagnosis and treatment of COVID-19 and revised via expert evaluation. It includes items on personal information; protection knowledge, skills, and behaviors related to COVID-19; access to information; and current information needs. The Cronbach alpha coefficient and the Kaiser-Meyer-Olkin value for the behavior section was 0.723 and 0.838, respectively, indicating that the research data were relatively true and reliable.

Personal information included demographic data, such as sex, age, educational level, occupation, and place of residence. The knowledge section was composed of 7 single-choice questions and 3 true-or-false questions, which were scored 3 points each, including initial symptoms, distribution of death cases, transmission routes, conditions for killing viruses, mask selection, medical observation days, fever temperature, new coronavirus infection after influenza vaccination, selection of disinfection products, and infectiousness of asymptomatic individuals. The skill section consisted of 9 single-choice questions, scored 3 points each, including cough or sneeze treatment, home quarantine measures, measures from outside to inside (ie, measures implemented by individuals when they

go back home from public places or workplaces), mask use, return notice (ie, matters that individuals should pay attention to when they come to or return to Jiangsu Province from other places), washing hands correctly, precautions, quarantine during travel, and attention to household alcohol disinfection. The behavior section comprised 11 scale questions, scored 0-3 points each, including no partying, wearing masks, wearing gloves, washing hands, no contact with live poultry, daily ventilation, weekly disinfection, distinction between the common cold and COVID-19, correct identification of epidemic information, workplace precautions, and community precautions. The highest possible score for each of these sections is 30, 27, and 33 points, respectively. The total score was calculated as follows:

> *Total protection score = (knowledge score + skill score + behavior score) / 90 × 100.*

Three methods were used to ensure data quality. Questionnaires filled out before 12 AM on February 15, 2020, were excluded, as the questionnaire was still in testing and was not officially published. Incomplete questionnaires were also excluded. Finally, questionnaires with irrelevant answers or obvious errors were excluded.

Statistical Analysis

Frequencies, proportions, arithmetic means, and standard deviations were used to present the data. The chi-square test, the independent samples *t* test (two-tailed), and a one-way analysis of variance were conducted, as appropriate. A multivariate linear regression analysis was performed to identify the factors associated with the knowledge score, skill score, behavior score, and total score for COVID-19. Further, a binary logistic regression analysis was used to explain the selection differences under different characteristics for key items. Unstandardized regression coefficients (β) and odds ratios and their 95% CIs were used to explain associations between variables. The questionnaire data were exported to Microsoft Excel 2016 (Microsoft Corp) for data processing and analysis in combination with SPSS 26.0 (IBM Corp). *P* values of <.05 were considered statistically significant.

In view of the impact of epidemic changes on public perceptions, the time-varying reproduction number (R_t) over a 7-day moving average and its 95% credible interval were estimated in R version 4.0.0 (R Foundation for Statistical Computing) using the method developed by Thompson et al [17], and the serial interval derived from a previous epidemiological survey [18], in combination with the officially published epidemic data of Jiangsu Province, mainland China, and the entire world. Thereafter, the maximal information coefficient [19,20] was applied to test for correlations among the daily participation number, average protection score, number of confirmed cases, and R_t.

Results

Participant Characteristics

In total, 52,066 participants were included in the investigation of the status of protection against COVID-19 from February 15 to April 21, 2020, after excluding 344 respondents (including 47 test accounts, 228 with incomplete answers, and 69 with irrelevant answers) (Table 1). Of these, there were 30,212 (58.03%) men, and the male-to-female sex ratio was 1.38:1. The study population mostly comprised those aged 31-40 years (19,131/52,066, 36.74%), followed by those aged 21-30 years (14,226/52,066, 27.32%) and 41-50 years (9885/52,066, 18.99%). In terms of educational level, the proportion of "junior college and bachelor's degree" was the largest, at 40.89% (21,291/52,066), and the proportion of "master's degree and above" was the smallest, at only 4.69% (2441/52,066). For the occupational classifications, enterprises (18,187/52,066, 34.93%) and business and service industries (5906/52,066, 11.34%) accounted for a relatively large proportion. One-third of the participants lived in rural areas, while the other two-thirds lived in urban areas. The number of participants involved in the investigation each day is shown in Figure 1, with 3 peaks, which generally corresponded to the time for resuming work and study in batches in Jiangsu Province.

Table 1. General characteristics and protection scores of participants.

Characteristic	Participants, n (%)	Total score Mean (SD)	Statistic	P value	Knowledge score Mean (SD)	Statistic	P value	Skill score Mean (SD)	Statistic	P value	Behavior score Mean (SD)	Statistic	P value
Sex			t=–12.4	<.001		t=–10.7	<.001		t=–11.2	<.001		t=–2.6	.009
Male	30,212 (58.03)	89.61 (8.99)			25.41 (4.23)			23.88 (4.14)			31.48 (2.89)		
Female	21,854 (41.97)	90.58 (8.66)			25.81 (4.19)			24.27 (3.84)			31.54 (2.78)		
Age (years)			F=387.1	<.001		F=245.0	<.001		F=381.2	<.001		F=34.7	<.001
≤20	5432 (10.43)	85.53 (11.74)			23.92 (5.42)			21.98 (5.30)			31.21 (3.44)		
21-30	14,226 (27.32)	90.76 (7.85)			25.99 (3.79)			24.42 (3.60)			31.38 (2.78)		
31-40	19,131 (36.74)	90.92 (7.90)			25.88 (3.93)			24.40 (3.61)			31.66 (2.63)		
41-50	9885 (18.99)	90.25 (8.71)			25.54 (4.18)			24.19 (3.89)			31.59 (2.80)		
51-60	3031 (5.82)	88.77 (10.13)			25.03 (4.51)			23.51 (4.58)			31.46 (3.05)		
≥61	361 (0.69)	85.14 (13.60)			23.65 (5.60)			22.34 (5.61)			30.79 (4.43)		
Educational level			F=1503.8	<.001		F=1320.0	<.001		F=1240.0	<.001		F=27.1	<.001
≤Junior high school	14,954 (28.72)	86.57 (10.54)			24.09 (4.90)			22.62 (4.82)			31.34 (3.27)		
High school and technical secondary school	13,380 (25.70)	89.51 (8.76)			25.22 (4.15)			23.83 (4.05)			31.62 (2.80)		
Junior college and bachelor's degree	21,291 (40.89)	92.45 (6.62)			26.67 (3.34)			25.06 (2.99)			31.55 (2.55)		
≥Master's degree	2441 (4.69)	92.83 (7.80)			27.06 (3.70)			25.12 (3.37)			31.42 (2.64)		
Occupation			F=282.4	<.001		F=233.6	<.001		F=224.0	<.001		F=19.4	<.001
Government agency and institution	3579 (6.87)	92.57 (7.66)			26.79 (3.57)			24.96 (3.42)			31.65 (2.56)		
Medical practitioner	2674 (5.14)	93.98 (7.36)			27.42 (3.35)			25.31 (3.26)			31.92 (2.57)		
Enterprise	18,187 (34.93)	91.26 (7.51)			26.11 (3.70)			24.63 (3.43)			31.49 (2.69)		
Business and service industry	5906 (11.34)	90.12 (8.08)			25.45 (4.00)			24.14 (3.74)			31.64 (2.67)		
Farmer[a]	2305 (4.43)	87.34 (10.44)			24.37 (4.81)			23.04 (4.69)			31.33 (3.30)		
Student	5507 (10.58)	87.13 (11.01)			24.52 (5.17)			22.67 (4.93)			31.35 (3.21)		
Freelancer	6282 (12.07)	88.62 (9.14)			24.85 (4.37)			23.47 (4.32)			31.57 (2.82)		

Characteristic	Participants, n (%)	Total score			Knowledge score			Skill score			Behavior score		
		Mean (SD)	Statistic	P value	Mean (SD)	Statistic	P value	Mean (SD)	Statistic	P value	Mean (SD)	Statistic	P value
Retiree	420 (0.81)	86.82 (11.36)			24.20 (5.02)			22.71 (4.96)			31.36 (3.02)		
Unemployed	1398 (2.69)	87.43 (10.37)			24.71 (4.68)			23.20 (4.44)			30.91 (3.52)		
Other	5808 (11.16)	88.80 (9.18)			25.03 (4.36)			23.58 (4.20)			31.43 (2.99)		
Place of residence			t=30.0	<.001		t=28.5	<.001		t=25.0	<.001		t=6.3	<.001
Urban area	34,426 (66.12)	90.90 (8.19)			25.97 (3.96)			24.37 (3.75)			31.56 (2.68)		
Rural area	17,640 (33.88)	88.31 (9.84)			24.81 (4.59)			23.40 (4.43)			31.39 (3.14)		
Total	52,066 (100.00)	90.02 (8.87)			25.58 (4.22)			24.05 (4.02)			31.51 (2.84)		

[a]"Farmer" includes agriculture, forestry, animal husbandry, sideline occupations, and fishery.

Figure 1. Daily number of participants and average total score.
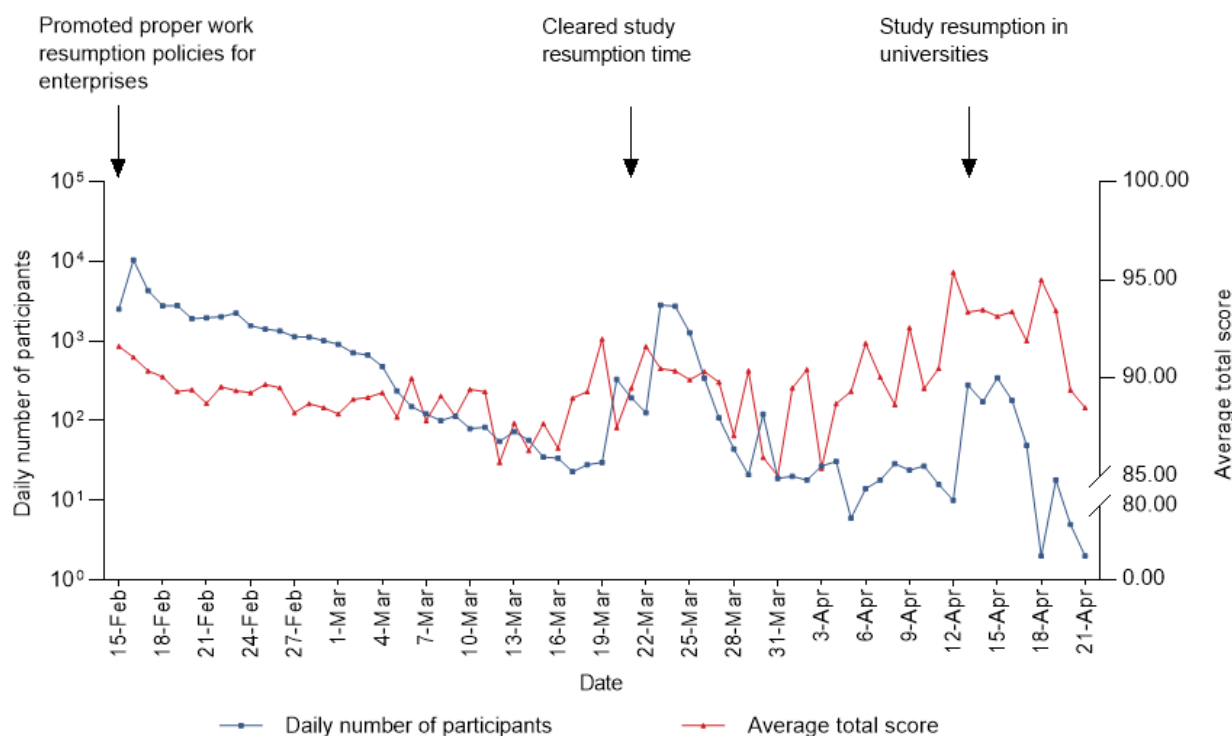
Protection Scores

For the total protection score, 65.91% of participants (34,315/52,066) had scores over 90 points. The univariate analysis showed that there were significant differences in the knowledge score, skill score, behavior score, and total score for sex, age, educational level, occupation, and place of residence (all Ps<.001, except for sex and behavior score [P=.009]; Table 1).

Partial Score

The protection score consisted of three parts: knowledge score, skill score, and behavior score. Initially, we analyzed the first two parts with the same scoring standard, with average scores of 25.58 (SD 4.22) and 24.05 (SD 4.02) (Table 1) and a range of correct answer rates of 62.01%-98.28% (Figure 2). The multivariate linear regression analysis indicated that women; those aged 21-60 years; those with an educational level of high school or greater; those with occupations categorized as government agency and institution, enterprise, business and

service industry, medical practitioners, and students; and those living in urban areas had significantly higher knowledge scores than men, those aged ≤20 years, those with an educational level of junior high school or less, those who were unemployed, and those who lived in rural areas (*P*<.001 or *P*=.007; Table S1 in Multimedia Appendix 2). Other than the above variables, those aged ≥61 years (*P*=.003), farmers (*P*=.01), and freelancers (*P*=.04) had significantly higher skill scores (Table S1 in Multimedia Appendix 2).

Figure 2. Rates of correct answers related to the knowledge and skill sections of the questionnaire. Reference line: 80.00%, shown in red.
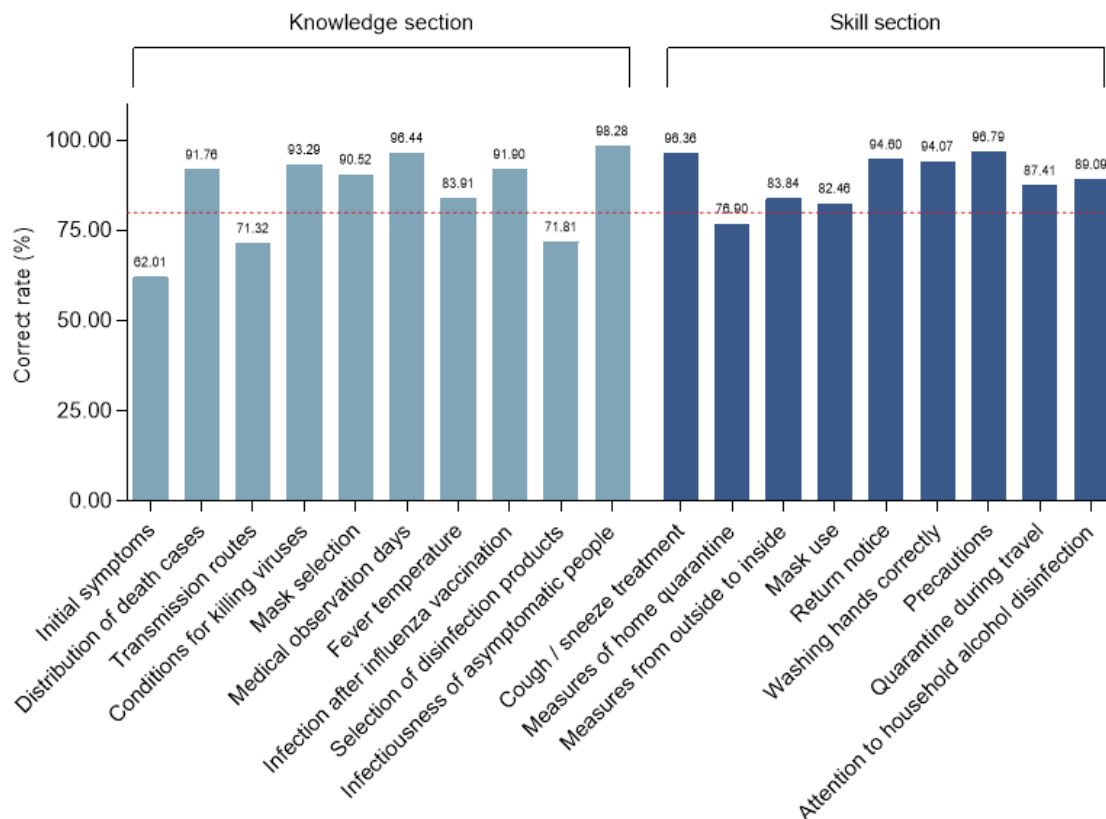
More concretely, the rates of correct answers for questions on medical observation days, infectiousness of asymptomatic individuals, cough or sneeze treatment, and precautions were higher than 95% in these two sections. Conversely, those of questions on initial symptoms (32,286/52,066, 62.01%), transmission routes (37,134/52,066, 71.32%), selection of disinfection products (37,390/52,066, 71.81%), and measures of home quarantine (40,037/52,066, 76.90%) were relatively low (Figure 2). There were significant differences in the answers to these four questions among the different sexes, age groups, educational levels, occupations, and places of residence (*P*<.001; Table S2 in Multimedia Appendix 2). The binary logistic regression analysis showed that the correct answer rates among women in relation to initial symptoms, transmission routes, selection of disinfection products, and measures of home quarantine were higher than those among men (*P*<.001; Table S3 in Multimedia Appendix 2). Those aged 21-40 years were more aware of disinfection products and measures of home quarantine than those aged ≤20 years (*P*<.001), while those aged ≥51 years were less aware of initial symptoms and transmission routes (*P*<.001 or *P*=.02). The correct rates among the participants with an educational level of high school or greater for these 4 questions were higher than those with an educational level of junior high school and below (*P*<.001). Those with occupations categorized under government agency

and institution and medical practitioners were more aware of the initial symptoms, transmission routes, disinfection products, and measures of home quarantine than those who were unemployed (*P*≤.001 or *P*=.002); in particular, medical practitioners had the highest correct answer rates. The correct answer rates for selection of disinfection products and home quarantine measures were higher among those with occupations categorized under enterprise (*P*=.02 or *P*=.001). Students had greater accuracy for initial symptoms, selection of disinfection products, and home quarantine measures (*P*<.001). Those living in urban areas had a higher accuracy for transmission routes and selection of disinfection products (*P*<.001 or *P*=.01; Table S3 in Multimedia Appendix 2).

Thereafter, the actual degree of protection among the participants was examined, and the average score was 31.51 (SD 2.84) (Tables 1 and 2). The multivariate linear regression analysis revealed that women (*P*=.01; Table S1 in Multimedia Appendix 2), those aged 21-60 years (*P*<.001), those with an educational level of high school and technical secondary school (*P*<.001) and junior college and bachelor's degree (*P*=.046), those with employment (*P*<.001), and those living in urban areas (*P*<.001) had significantly higher behavior scores than men, those aged ≤20 years, those with an educational level of

junior high school or below, those who were unemployed, and those who lived in rural areas.

Table 2. Protection behaviors and the degree to which participants were able to implement these behaviors (as indicated by the 3-point response "able to").

Behavior	Participants, n (%)
No partying	48,955 (94.02)
Wearing masks	50,989 (97.93)
Wearing gloves	46,607 (89.52)
Washing hands	49,607 (95.28)
No contact with live poultry	50,191 (96.40)
Daily ventilation	50,670 (97.32)
Weekly disinfection	41,776 (80.24)
Distinction between the common cold and COVID-19	42,477 (81.58)
Correct identification of epidemic information	50,908 (97.78)
Workplace precautions	46,800 (89.89)
Community precautions	47,009 (90.29)

Specifically, a higher proportion of participants were able to avoid gatherings, wear gloves, wash hands in a timely manner, keep away from live poultry and livestock, ventilate each day, and identify information related to the epidemic correctly and believed that precautions in workplaces or communities were in place. For example, 97.93% (50,989/52,066) of participants could wear masks properly when they went out (Table 2). However, 19.76% (10,290/52,066) still could not disinfect their homes each week, which was significantly associated with age, educational level, occupation, and place of residence ($P<.001$; Table S2 in Multimedia Appendix 2). Similarly, 18.42% (9589/52,066) could not distinguish the initial symptoms of the common cold and COVID-19, and this was significantly related to sex, age, educational level, occupation, and place of residence ($P<.001$; Table S2 in Multimedia Appendix 2). The binary logistic regression analysis indicated that compared with participants aged ≤20 years, those aged 31-60 years could disinfect their homes weekly ($P<.001$; Table S3 in Multimedia Appendix 2), and those aged 41-50 years were aware of initial symptom differences between the common cold and COVID-19 ($P=.04$). Those with a high educational level were unable to disinfect their homes weekly and clearly distinguish between the common cold and COVID-19 ($P<.001$ or $P=.02$). In addition to the retirees ($P=.08$), those with the other indicated occupations were able to disinfect their homes weekly ($P<.001$ or $P=.003$).

With the exception of those with occupations categorized under enterprise ($P=.06$), people in the other profession categories could distinguish between the common cold and COVID-19 (all Ps<.05). Participants living in urban areas were more often able to disinfect their homes weekly than those living in rural areas ($P=.04$; Table S3 in Multimedia Appendix 2).

Total Score

The total score was obtained by summing up the scores of the three abovementioned parts and converted to the hundred-mark system. The average total protection score was 90.02 (SD 8.87), rising with fluctuations over time (Figure 1), with the highest score (mean 93.98, SD 7.36) observed among medical practitioners (Table 1). The five demographic characteristics (ie, sex, age, educational level, occupation, and place of residence) significant in the univariate analyses ($P<.001$; Table 1) were included in the multivariate linear regression analysis ($F_{19,52046}=343.426$, $P<.001$). This showed that women, those aged 21-60 years, those with an educational level of high school or above, those with occupations other than being a retiree, and those living in urban areas had significantly higher total protection scores than men, those aged ≤20 years, those with an educational level of junior high school and below, those who were unemployed, and those who lived in rural areas ($P<.001$; Table 3).

Table 3. Results of the multivariate linear regression analysis on factors influencing the total protection score.

Variable	Coefficient	SE	95% CI	t test	P value	Collinearity statistics (VIF[a])
Constant	79.914	0.311	79.305 to 80.523	257.238	<.001	N/A[b]
Sex (reference: male)						
Female	0.971	0.077	0.820 to 1.123	12.576	<.001	1.082
Age (years; reference: ≤ 20)						
21-30	4.592	0.221	4.159 to 5.026	20.778	<.001	7.224
31-40	5.481	0.226	5.039 to 5.924	24.282	<.001	8.821
41-50	4.963	0.232	4.508 to 5.419	21.363	<.001	6.183
51-60	3.252	0.267	2.728 to 3.776	12.164	<.001	2.919
≥61	0.199	0.535	–0.849 to 1.248	0.373	.71	1.467
Educational level (reference: ≤junior high school)						
High school and technical secondary school	2.200	0.103	1.997 to 2.402	21.312	<.001	1.515
Junior college and bachelor's degree	4.206	0.105	4.000 to 4.412	40.056	<.001	1.985
≥Master's degree	4.312	0.197	3.925 to 4.698	21.851	<.001	1.296
Occupation (reference: unemployed)						
Government agency and institution	2.712	0.272	2.178 to 3.245	9.964	<.001	3.531
Medical practitioner	4.567	0.281	4.016 to 5.119	16.232	<.001	2.872
Enterprise	2.272	0.237	1.807 to 2.737	9.580	<.001	9.525
Business and service industry	1.762	0.251	1.271 to 2.254	7.027	<.001	4.710
Farmer[c]	1.164	0.287	0.601 to 1.727	4.055	<.001	2.598
Student	3.830	0.304	3.235 to 4.426	12.609	<.001	6.500
Freelancer	1.078	0.248	0.591 to 1.565	4.338	<.001	4.879
Retiree	0.647	0.513	–0.359 to 1.653	1.261	.21	1.569
Other	0.993	0.250	0.503 to 1.483	3.970	<.001	4.613
Place of residence (reference: rural area)						
Urban area	1.046	0.083	0.884 to 1.209	12.623	<.001	1.146

[a]VIF: variance inflation factor.

[b]N/A: not applicable.

[c]"Farmer" included agriculture, forestry, animal husbandry, sideline occupations, and fishery.

XSL•FO
RenderX

Information Acquisition and Information Needs

In this study, access to personal protection knowledge on COVID-19 and information needs were also investigated. Television shows, government websites, and news outlets (46,145/52,066, 88.63%), as well as the government's WeChat public account (45,657/52,066, 87.69%), were the main sources for acquiring personal protection knowledge. The sources for participants of different sexes and places of residence were similar. Participants of all ages rarely obtained protection information from microblogs or via communication among family, relatives, and friends. Those with an educational level of high school or less obtained information mostly from television shows, government websites, and news outlets, while those with an educational level of junior college or above obtained information from the government's WeChat public account. Participants with occupations categorized under government agency, institution, and enterprise, as well as medical practitioners, obtained information more often from the government's WeChat public account. Participants believed that the government's media and WeChat public accounts (48,307/52,066, 92.78%) and authoritative medical experts (46,062/52,066, 88.47%) were the most reliable information sources (Table S4 in Multimedia Appendix 2).

The current information needs of the participants included the latest epidemic developments (46,729/52,066, 89.75%), disease treatment progress (42,181/52,066, 81.01%), and daily protection knowledge (41,451/52,066, 79.61%). Participants of different sexes had large differences in information needs in terms of disease treatment progress, prevention and control status in epidemic areas, and social dynamics; the differences in the other aspects were smaller. Participants of different ages, especially those aged 21-60 years, were very eager to understand the latest epidemic developments. The information needs of those with an educational level of junior high school or below concerned the latest epidemic developments and daily protection knowledge; conversely, the information needs of those with an educational level of high school or above were the latest epidemic developments and disease treatment progress. All participants paid less attention to material supply and social dynamics. Those with occupations categorized under government agency and institution, enterprise, business and service industry, freelancers, medical practitioners, and those who were unemployed had higher needs for epidemic developments and disease treatment progress, while farmers, students, and retirees had higher needs for epidemic developments and protection knowledge. Conversely, those living in rural areas were more interested in obtaining information on epidemic developments and daily protection knowledge than those living in urban areas (Table S5 in Multimedia Appendix 2).

Correlation Among the Daily Participation Number, Average Protection Score, Number of Confirmed Cases, and R_t

We analyzed the R_t trends and attempted to determine associations among the daily participation number, average protection score, number of confirmed cases, and R_t values during the investigation. Owing to the implementation of strict control measures [21], the R_t in Jiangsu Province declined below 1, close to 0, and the R_t in mainland China also dropped significantly (Figures 3A and B). However, the number of confirmed cases worldwide has been increasing, with the global R_t showing a trend of first rising and then declining and maintaining a value around 1 (Figure 3C). The correlation analysis revealed that the daily participation number and average protection score were significantly associated with the number of confirmed cases and R_t in Jiangsu Province, mainland China, and the entire world (maximal information coefficient >0.70, range: 0.76-1.00; Figure 4).

Figure 3. Time-varying reproduction numbers (R_t), their 95% CIs, and confirmed cases for Jiangsu Province, mainland China, and the entire world.
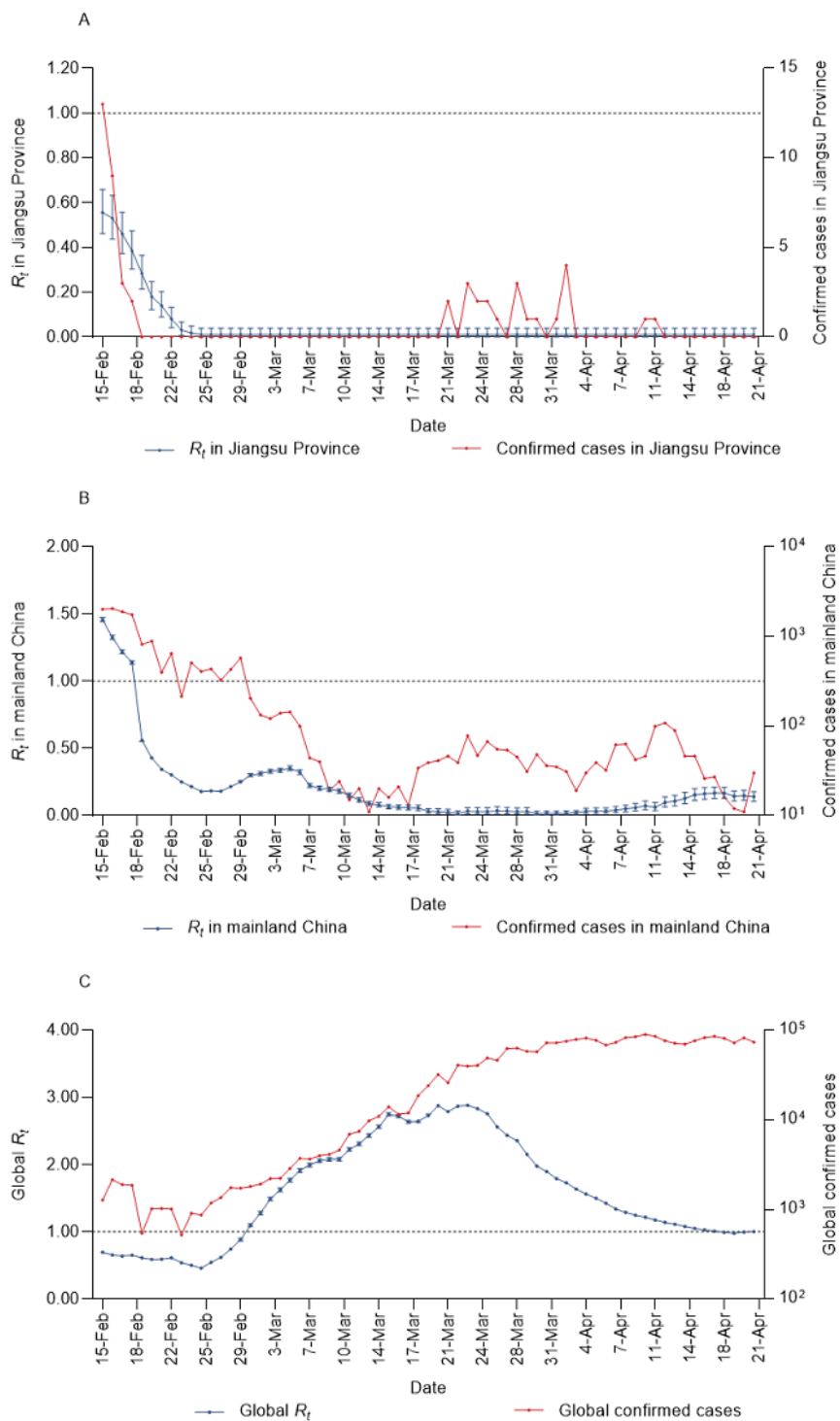
Figure 4. Correlation results for the daily participation number, average protection score, confirmed cases, and time-varying reproduction number (R_t).
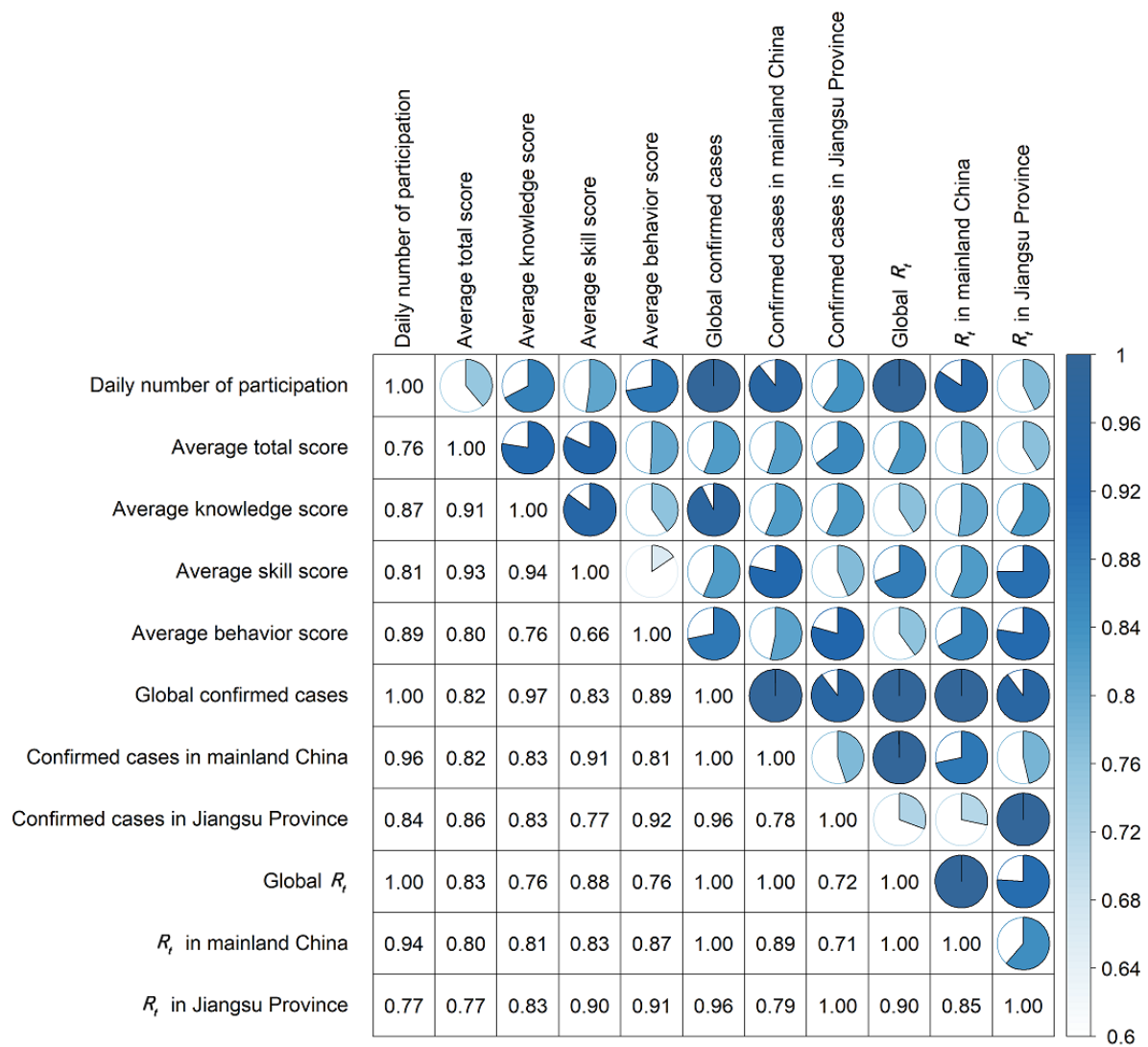
Discussion

Principal Results and Comparison With Prior Work

Emerging infectious diseases are usually unpredictable with a lack of effective vaccines and drug treatments, and have direct or indirect negative impacts on economic development, social stability, and the public's quality of life [22,23]. During the prior outbreaks of SARS and MERS, investigations were conducted to understand the public's knowledge, attitudes, and practices [24-29]. Similarly, this study has shown that there was a relatively strong relationship between epidemic development and public perceptions.

A total of 52,066 participants were included in the study, of whom 65.91% had a total protection score above 90 points, indicating that the protection knowledge and skills were well understood and the actual action ability was strong. Unfortunately, there were still deficiencies in knowledge, skills, and actual behaviors in 34.09% of participants, and, hence, precise health education measures need to be provided.

For the knowledge section, the correct answer rates for initial symptoms, transmission routes, and selection of disinfection products were less than 80%; specifically, the rate for initial symptoms was only 62.01% (32,286/52,066), which was less than that of hospital visitors [14]. This may be because the outbreak time of COVID-19 overlaps with that of the common cold, flu, and other diseases, and these are also respiratory diseases, which have certain similarities in clinical manifestations. Similarly, owing to limited knowledge on COVID-19, which is currently an emerging infectious disease, individuals will tend to choose transmission routes and disinfection products that they consider reasonable. It is suggested that relevant departments should release authoritative health information on COVID-19 in a timely manner and strengthen the promotion and education of daily protection knowledge to meet the public's needs.

For the skill section, the correct rate for home quarantine measures was relatively low; 23.10% (12,029/52,066) of the respondents were unaware that they cannot participate in family dinners. Yet, it is necessary to ensure that dishes and chopsticks

are used and sterilized separately to avoid cross-infection. This may be attributed to the fact that although most quarantined individuals have no symptoms or mild symptoms, there is still a probability of presymptomatic transmission [30], which prompts relevant departments to provide health tips on epidemic prevention and control and consider such families as key monitoring objects.

For the actual behavior section, most individuals could reduce risk behaviors and take necessary protective measures, similar with previous study findings [11,13,31]. However, the proportion of those able to disinfect their homes weekly and distinguish the initial symptoms between the common cold and COVID-19 was lower than that of those presenting other behaviors, as in a previous study [14]. This may be related to the limited knowledge on COVID-19 and the lack of self-protection ability, indicating that public health information literacy needs to be improved. In the regression analysis, those with high educational levels were relatively unable to disinfect their homes weekly and clearly distinguish between the common cold and COVID-19. The reason may be that these individuals usually pay more attention to personal protection and that they are more cautious in answering the questions on the difference between the common cold and COVID-19.

Sex, age, educational level, occupation, and place of residence affected the total protection score for COVID-19 at different degrees. Women (average score of 90.58, SD 8.66) tended to have higher total protection scores than men, which is similar to the findings of previous investigations [13,31]. A large difference between men and women was observed in the skill scores—23.88 (SD 4.14) for men and 24.27 (SD 3.84) for women. Conversely, the total protection scores of those aged 21-60 years tended to be higher than the scores of those aged ≤20 years, which may be attributed to the current situation of resuming work and study; thus, these individuals need to actively obtain information on protection knowledge for COVID-19 and improve their self-protection ability. The total score was influenced by the educational level; the total score of those with higher educational levels tended to be higher than that of those with lower educational levels. Relevant studies also showed that individuals with higher educational levels were more willing to accept new knowledge and skills and adopt healthier practices [14,32,33]. In comparison with the unemployed, all participants, except the retirees and especially those with occupations categorized under government agency and institution, enterprise, business and service industry, medical practitioners, and students were more likely to have higher total scores, which may be attributed to their professional characteristics. Those living in urban areas had higher total protection scores than those living in rural areas. This may be attributed to the insufficient basic medical resources and relatively weak primary public health prevention strategies in

rural areas. Further, the results may be associated with the relatively limited access to the internet and online health information resources [13]. Therefore, it is necessary to focus on the dissemination of health education knowledge on COVID-19 for men, those aged ≤20 years, those with an educational level of junior high school or less, those who are unemployed, and those living in rural areas.

This study showed that the government, television shows, and news outlets were the main sources for protection knowledge, which accounted for a higher proportion than that in a US sample [12]. Information released by the government and authoritative medical experts was also considered as reliable information, which is different from that reported in a previous study [31]. Moreover, this study emphasizes the need to continue to publicize the latest epidemic developments, disease treatment progress, daily protection knowledge, and other information on COVID-19.

Limitations

There were several limitations in this study. First, the questionnaire used was designed based on a literature review and was used for the investigation after being revised via expert consultations. With the absence of a rigorous design process, the reliability of information may decline. Second, since the online surveys were conducted through WeChat and only in Jiangsu Province, the research samples are biased and limited. Given the differences in the resumption of work and study across different regions, our conclusions may change when expanding the study population. Lastly, the questionnaire only considered the influencing factors of the total protection score for COVID-19 from an individual perspective, without considering the influence of macro factors, such as government policies and society.

Conclusion

A high proportion of study participants had good protection knowledge and skills related to COVID-19. The factors influencing the total protection score for COVID-19 included sex, age, educational level, occupation, and place of residence. The study results suggest that relevant government departments need to update accurate information on COVID-19 in a timely manner, such as the latest epidemic developments and disease treatment progress, via official media and new media channels, and continue to promote daily protection knowledge on COVID-19. When resuming work and study, relevant departments need to apply different health education measures and conduct extensive and in-depth health education and promotion activities to guide the public, especially men, younger individuals, individuals with low educational levels, the unemployed, and individuals living in rural areas, to adopt positive and healthy behaviors. Doing so will ultimately reduce the negative impact of COVID-19.

XSL·FO

RenderX

Authors' Contributions

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[PDF File (Adobe PDF File), 238 KB - jmir_v22i11e21672_app1.pdf]

Multimedia Appendix 2
Supplementary tables.
[PDF File (Adobe PDF File), 533 KB - jmir_v22i11e21672_app2.pdf]

References

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
3. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization. 2020 Jan 30. URL: https://tinyurl.com/yxns3jfx [accessed 2020-05-30]
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
5. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. N Engl J Med 2020 Feb 20;382(8):692-694. [doi: 10.1056/NEJMp2000929] [Medline: 31978293]
6. Coronavirus disease 2019 (COVID-19) Situation Report – 27. World Health Organization. 2020 Feb 16. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200216-sitrep-27-covid-19.pdf?sfvrsn=78c0eb78_4 [accessed 2020-05-30]
7. Coronavirus disease 2019 (COVID-19) Situation Report – 93. World Health Organization. 2020 Apr 22. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200422-sitrep-93-covid-19.pdf?sfvrsn=35cf80d7_4 [accessed 2020-05-30]
8. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 May;19(5):305-306. [doi: 10.1038/d41573-020-00073-5] [Medline: 32273591]
9. Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Cent Sci 2020 Mar 25;6(3):315-331 [FREE Full text] [doi: 10.1021/acscentsci.0c00272] [Medline: 32226821]
10. Lohiniva A, Sane J, Sibenberg K, Puumalainen T, Salminen M. Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020. Euro Surveill 2020 Apr;25(13) [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000317] [Medline: 32265008]
11. Geldsetzer P. Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
12. McFadden SM, Malik AA, Aguolu OG, Willebrand KS, Omer SB. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
13. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
14. Kebede Y, Yitayih Y, Birhanu Z, Mekonen S, Ambelu A. Knowledge, perceptions and preventive practices towards COVID-19 early in the outbreak among Jimma university medical center visitors, Southwest Ethiopia. PLoS One 2020;15(5):e0233744 [FREE Full text] [doi: 10.1371/journal.pone.0233744] [Medline: 32437432]

15. Pérez-Fuentes MDC, Molero Jurado MDM, Oropesa Ruiz NF, Martos Martínez Á, Simón Márquez MDM, Herrera-Peco I, et al. Questionnaire on Perception of Threat from COVID-19. J Clin Med 2020 Apr 22;9(4):1196 [FREE Full text] [doi: 10.3390/jcm9041196] [Medline: 32331246]

16. Jiangsu health code officially launched!. Jiangsu Provincial Center for Disease Control and Prevention. 2020 Mar 5. URL: http://www.jscdc.cn/zxzx/jkzt1/tfyjzt/xxgzbd/skm/202003/t20200305_68138.html [accessed 2020-05-30]

17. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics 2019 Dec;29:100356 [FREE Full text] [doi: 10.1016/j.epidem.2019.100356] [Medline: 31624039]

18. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]

19. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al. Detecting novel associations in large data sets. Science 2011 Dec 16;334(6062):1518-1524 [FREE Full text] [doi: 10.1126/science.1205438] [Medline: 22174245]

20. Albanese D, Filosi M, Visintainer R, Riccadonna S, Jurman G, Furlanello C. Minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. Bioinformatics 2013 Feb 01;29(3):407-408. [doi: 10.1093/bioinformatics/bts707] [Medline: 23242262]

21. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis 2020 Jul;20(7):793-802 [FREE Full text] [doi: 10.1016/S1473-3099(20)30230-9] [Medline: 32247326]

22. Howard CR, Fletcher NF. Emerging virus diseases: can we ever expect the unexpected? Emerg Microbes Infect 2012 Dec;1(12):e46 [FREE Full text] [doi: 10.1038/emi.2012.47] [Medline: 26038413]

23. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature 2008 Feb 21;451(7181):990-993 [FREE Full text] [doi: 10.1038/nature06536] [Medline: 18288193]

24. Brug J, Aro AR, Oenema A, de Zwart O, Richardus JH, Bishop GD. SARS risk perception, knowledge, precautions, and information sources, the Netherlands. Emerg Infect Dis 2004 Aug;10(8):1486-1489 [FREE Full text] [doi: 10.3201/eid1008.040283] [Medline: 15496256]

25. Lau JTF, Yang X, Pang E, Tsui HY, Wong E, Wing YK. SARS-related perceptions in Hong Kong. Emerg Infect Dis 2005 Mar;11(3):417-424 [FREE Full text] [doi: 10.3201/eid1103.040675] [Medline: 15757557]

26. Koh D, Lim MK, Chia SE, Ko SM, Qian F, Ng V, et al. Risk perception and impact of Severe Acute Respiratory Syndrome (SARS) on work and personal lives of healthcare workers in Singapore: what can we learn? Med Care 2005 Jul;43(7):676-682. [doi: 10.1097/01.mlr.0000167181.36730.cc] [Medline: 15970782]

27. Sahin MK, Aker S, Kaynar Tuncel E. Knowledge, attitudes and practices concerning Middle East respiratory syndrome among Umrah and Hajj pilgrims in Samsun, Turkey, 2015. Euro Surveill 2015;20(38) [FREE Full text] [doi: 10.2807/1560-7917.ES.2015.20.38.30023] [Medline: 26535787]

28. Al-Hazmi A, Gosadi I, Somily A, Alsubaie S, Bin Saeed A. Knowledge, attitude and practice of secondary schools and university students toward Middle East Respiratory Syndrome epidemic in Saudi Arabia: A cross-sectional study. Saudi J Biol Sci 2018 Mar;25(3):572-577 [FREE Full text] [doi: 10.1016/j.sjbs.2016.01.032] [Medline: 29686521]

29. Alqahtani AS, Wiley KE, Mushta SM, Yamazaki K, BinDhim NF, Heywood AE, et al. Association between Australian Hajj Pilgrims' awareness of MERS-CoV, and their compliance with preventive measures and exposure to camels. J Travel Med 2016 May;23(5):taw046 [FREE Full text] [doi: 10.1093/jtm/taw046] [Medline: 27432904]

30. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020 May;26(5):672-675. [doi: 10.1038/s41591-020-0869-5] [Medline: 32296168]

31. Huang Y, Wu Q, Wang P, Xu Y, Wang L, Zhao Y, et al. Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study. J Med Internet Res 2020 May 12;22(5):e18718 [FREE Full text] [doi: 10.2196/18718] [Medline: 32396516]

32. Oh J, Jeong BY, Yun EH, Lim MK. Awareness of and Attitudes toward Human Papillomavirus Vaccination among Adults in Korea: 9-Year Changes in Nationwide Surveys. Cancer Res Treat 2018 Apr;50(2):436-444 [FREE Full text] [doi: 10.4143/crt.2017.174] [Medline: 28494533]

33. Rossi MDSC, Stedefeldt E, da Cunha DT, de Rosso VV. Food safety knowledge, optimistic bias and risk perception among food handlers in institutional food services. Food Control 2017 Mar;73:681-688. [doi: 10.1016/j.foodcont.2016.09.016]

Abbreviations

BIP-Q5: Brief Illness Perception Questionnaire
Jiangsu Provincial CDC: Jiangsu Provincial Center for Disease Control and Prevention
MERS: Middle East respiratory syndrome
R_t: time-varying reproduction number
SARS: severe acute respiratory syndrome

XSL•FO
RenderX

WHO: World Health Organization

Original Paper

Public Emotions and Rumors Spread During the COVID-19 Epidemic in China: Web-Based Correlation Study

Wei Dong[1*], PhD; Jinhu Tao[1*], MA; Xiaolin Xia[1*], BA; Lin Ye[2*], MA; Hanli Xu[3], PhD; Peiye Jiang[4], PhD; Yangyang Liu[1], PhD

[1]School of Education, Tianjin University, Tianjin, China

[2]School of Media and Communication, Shanghai Jiaotong University, Shanghai, China

[3]College of Life Sciences and Bioengineering, School of Science, Beijing Jiaotong University, Beijing, China

[4]Office of International Cooperation and Exchanges, Nanjing University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Yangyang Liu, PhD
School of Education
Tianjin University
135 Yaguan Road
Tianjin, 300350
China
Phone: 86 18118849345
Email: liuyangyang661@sina.com

Abstract

Background: Various online rumors have led to inappropriate behaviors among the public in response to the COVID-19 epidemic in China. These rumors adversely affect people's physical and mental health. Therefore, a better understanding of the relationship between public emotions and rumors during the epidemic may help generate useful strategies for guiding public emotions and dispelling rumors.

Objective: This study aimed to explore whether public emotions are related to the dissemination of online rumors in the context of COVID-19.

Methods: We used the web-crawling tool Scrapy to gather data published by People's Daily on Sina Weibo, a popular social media platform in China, after January 8, 2020. Netizens' comments under each Weibo post were collected. Nearly 1 million comments thus collected were divided into 5 categories: happiness, sadness, anger, fear, and neutral, based on the underlying emotional information identified and extracted from the comments by using a manual identification process. Data on rumors spread online were collected through Tencent's Jiaozhen platform. Time-lagged cross-correlation analyses were performed to examine the relationship between public emotions and rumors.

Results: Our results indicated that the angrier the public felt, the more rumors there would likely be ($r=0.48$, $P<.001$). Similar results were observed for the relationship between fear and rumors ($r=0.51$, $P<.001$) and between sadness and rumors ($r=0.47$, $P<.001$). Furthermore, we found a positive correlation between happiness and rumors, with happiness lagging the emergence of rumors by 1 day ($r=0.56$, $P<.001$). In addition, our data showed a significant positive correlation between fear and fearful rumors ($r=0.34$, $P=.02$).

Conclusions: Our findings confirm that public emotions are related to the rumors spread online in the context of COVID-19 in China. Moreover, these findings provide several suggestions, such as the use of web-based monitoring methods, for relevant authorities and policy makers to guide public emotions and behavior during this public health emergency.

(J Med Internet Res 2020;22(11):e21933)   doi:10.2196/21933

KEYWORDS

Introduction

In December 2019, the earliest cases of COVID-19 were reported in Wuhan, Hubei Province, China. On January 23, 2020, several cities in Hubei Province were quarantined in an attempt to slow down community transmission of the disease. On January 30, 2020, the World Health Organization officially announced that the COVID-19 outbreak was a public health emergency of international concern [1]. As COVID-19 has become a serious global problem, people all over the world are faced with the risk of infection; this has caused a change in people's behaviors as well as fluctuations in their emotions [2]. During COVID-19 outbreaks in different parts of the world, there was a proliferation of emotional comments and occasional rumors posted on the internet. Although a number of epidemiological and clinical studies have been performed thus far, few studies have examined public emotional response, and to our knowledge, no study has investigated whether public emotions are related to the occasional online rumors spread in the context of COVID-19. In this study, we seek to contribute to the literature by addressing this gap in the existing research.

Rumors have abounded since ancient times, but they have been the focus of researchers only since the Second World War [3]. In a pioneer study, Prasad [4] proposed that rumors are spread due to the anxiety and fear of unknown things in a disaster. Allport and Postman [5] suggested that the spread of rumors is a process of releasing emotions by telling stories to others. The lack of barriers of the internet enables individuals to participate in online discussions anonymously, which creates a favorable environment for the spread of rumors [6]. Some empirical studies have suggested that among the messages forwarded on the internet by users, those with emotional elements (anger, sadness, or anxiety) were 36.7% more frequent than those without emotional elements [7] and that people are more likely to believe online rumors with the same type of emotion that they currently feel. That is, angry people are more receptive to anger-related rumors [8]. However, until now, online rumors have not been extensively researched. All prior research has focused on analyzing rumors generated due to anxiety and fear. It is unclear whether other emotions are related to the spread of rumors. Additionally, no empirical studies have examined whether there is a significant correlation between emotions and rumors in the context of major public health emergencies.

In the context of COVID-19, various online rumors have led to inappropriate behaviors among people in response to the epidemic, which have adversely affected people's physical and mental health [9]. We believe that the findings of this study can provide some useful strategies to guide public emotions and dismiss rumors, in an effort to fight this global crisis.

Previous studies on public emotions during major public health emergencies have found that people usually experience negative emotions such as panic, anxiety, anger, grief, and disgust [10,11], likely due to the illnesses and deaths that occur during these emergencies. For example, 17 years ago, a study on SARS found that tension, helplessness, panic, and anger were significantly related to an increase in the number of SARS cases [12]. Some other studies also showed that public anxiety

changed with the development of the epidemic (eg, avian influenza A [H7N9] or influenza A [H1N1]) [13,14]. In line with some previous studies that used data from different sources [15,16], we aimed to examine the dynamic relationship between the public's emotions and rumors in the context of COVID-19, by using data from China's mainstream internet media platforms (Sina Weibo and Tencent). According to a report released by the China Internet Network Information Center, as of March 2020, there were 904 million internet users and the internet penetration rate in China had reached 64.5% [17]. Hence, the data obtained through the online responses from netizens are considered to be representative of the public's response to major health emergencies [18]. Moreover, social media surveillance can provide real-time information on the public's response; solve the problem of underrepresentation of samples; and avoid recall or reporting bias caused by personal observation, collection, and recording [19,20]. According to cognitive dissonance theory [21], rumors may serve as a channel for the public to reduce or eliminate cognitive dissonance caused by emotions. In this study, we aimed to explore whether the public's emotions are related to the generation and dissemination of rumors in the context of COVID-19.

Methods

Epidemiological Data

On January 20, 2020, China launched a monitoring and quarantine system to record daily information about COVID-19 cases. The epidemiological data, issued by the National Health Commission of People's Republic of China (NHC), were collected at the grassroots level on a daily basis. For this study, we used the epidemiological data (specifically, the daily number of newly confirmed COVID-19 cases) from all provincial-level regions of China that were released by the NHC from January 20 to March 10, 2020.

Data on Public Emotions

Sina Weibo, the Chinese version of Twitter, is the largest social media platform in China. People's Daily, with more than 116 million followers on Weibo, is one of the most influential and authoritative news media on Sina Weibo. In this study, we used the web crawler Scrapy to gather relevant data published by People's Daily on Weibo between January 20 and March 8, 2020, as well as netizens' comments under each Weibo post. Millions of real-time comment texts contain substantial emotional information; we identified and extracted these public emotions from the comments collected.

Although many sentiment analysis tools based on natural language processing or machine learning [22,23] are able to automatically extract emotions from the text by utilizing less time and labor, the accuracy of machine identification is lower than that of manual identification. Some simple classifications, such as positive or negative, can be detected through machine identification; however, for more subtle discrete emotions, such as anger, fear, and happiness, manual identification would yield better accuracy. Thus, we adopted manual identification methods in this study. Recently, Jack et al [24] reported that people universally experience only 4 basic emotions: happiness, sadness, fear, and anger. In this study, public emotions were

divided into 5 categories: the 4 abovementioned basic human emotions (ie, happiness, sadness, fear, and anger) and "neutral" (ie, representing no emotions). A total of 143 trained research assistants participated in the manual identification process. Each online comment was identified and classified by at least 3 trained research assistants, and conflicting classification results were resolved by a majority consensus to arrive at the final classification. The interrater reliability of the coders was "substantial" (κ=0.77) [25].

Data on Emotional Rumors

Tencent's Jiaozhen [26] is the first network-wide, professional, and timely fact-checking platform in China that checks internet news that has been widely circulated and dispels rumors on a daily basis. Data on online rumors about COVID-19 disseminated between January 20 and March 10, 2020, were collected through this platform, and manual identification was performed to identify all kinds of emotional rumors spread during this period.

Data Calibration and Statistical Analysis

Figure 1A shows that the daily number of newly confirmed cases in Hubei Province suddenly surged on February 12, 2020, when the diagnosis criteria were revised. Thereafter, the COVID-19 diagnosis criteria used in Hubei was consistent with those outside the province. Therefore, the epidemiological data of newly confirmed cases in Hubei that were collected before February 14 were calibrated before the follow-up analysis.

Figure 1. (A) Daily confirmed cases of COVID-19 in China. Blue dots represent numbers for the whole country, red triangles represent numbers for Hubei Province, and black circles represent numbers for other provinces. (B) Calibration of the daily number of confirmed cases in Hubei Province. Red triangles represent the daily number of confirmed cases in the province before calibration, and black dots represent the daily number of confirmed cases in the province after calibration. The sharp peak disappeared after calibration. HB: Hubei Province.
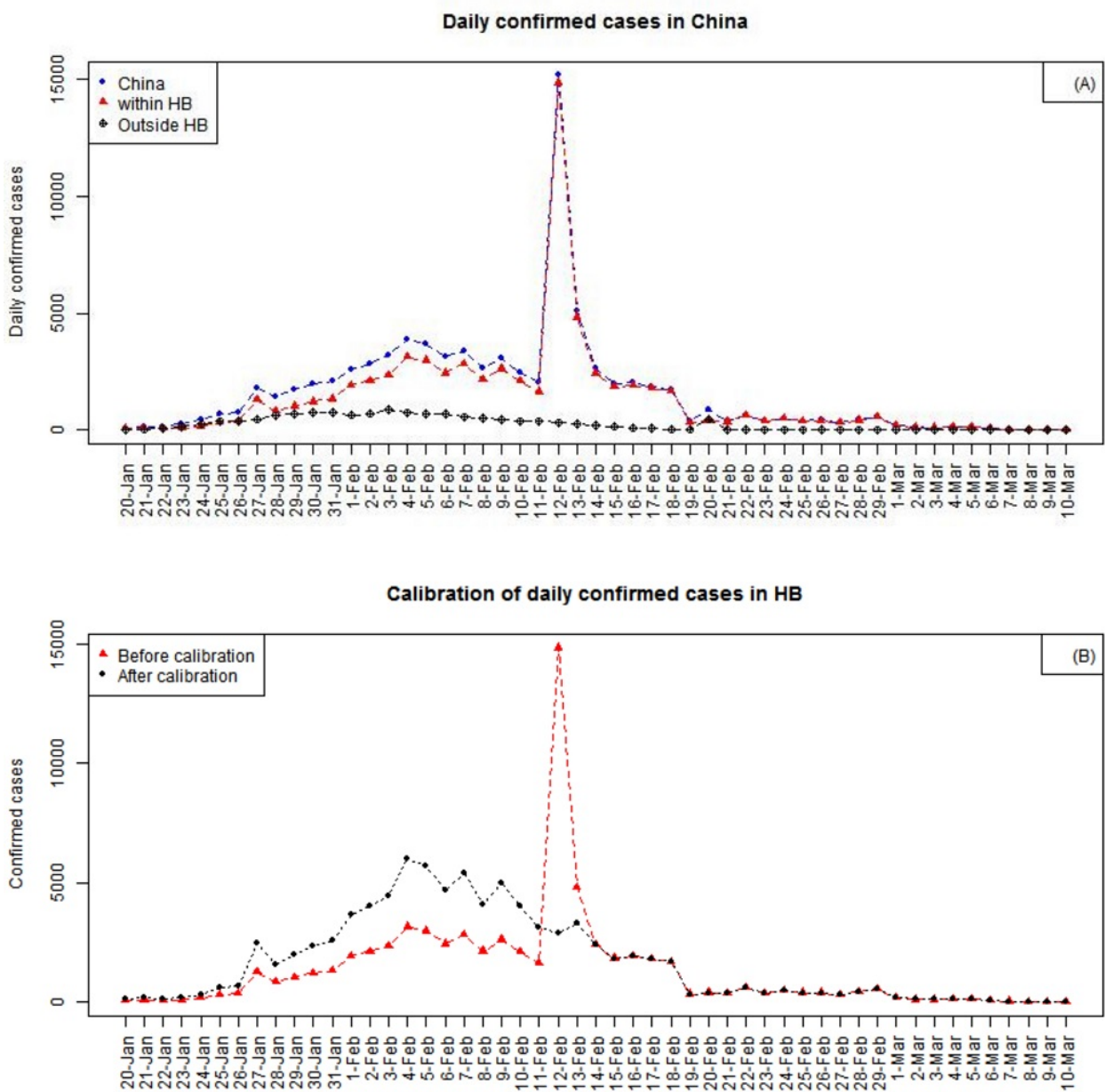
According to the Hubei Health Committee, on February 12 and 13, 2020, the ratios of COVID-19 cases that were clinically diagnosed and those that were detected based on nucleic acid testing were 8.44 (13,332/1580) and 1.79 (3095/1728), respectively. Most of these cases were a result of long-term accumulation of suspected cases. The decline in the number of clinically diagnosed cases on February 13, 2020, suggested that the cumulative suspected cases had been processed within those 2 days. The corresponding ratios of COVID-19 cases that were clinically diagnosed and those detected via nucleic acid testing on February 14 and 15, 2020, were 0.89 (1138/1282) and 0.93 (888/955), which average at 0.91. Therefore, from January 20 to February 13, 2020, the calibrated daily number of newly confirmed cases (as shown in Figure 1B) was the number of cases confirmed by nucleic acid detection multiplied by (1+0.91).

As the data on daily epidemic situations and emotional comments published on Weibo showed an exponential distribution, the data were log-transformed to stabilize the variance of time-series before conducting further analysis.

We used time-lagged cross-correlation to examine the relationship between public emotions and the total number of rumors and that between public emotions and the rumors with different types of emotions. We also used Pearson correlation coefficient to examine these relationships, with the maximum time lag of 10 days and a P value smaller than .05 as a threshold.

Results

Rumors in the Context of the COVID-19 Epidemic

Among the 276 rumors collected, 176 (63.8%) were neutral rumors, 62 (22.5%) were fear-related rumors, 19 (6.9%) were happiness-related rumors, 12 (4.4%) were anger-related rumors, and 7 (2.5%) were sadness-related rumors. Figure 2A shows the relationship between the COVID-19 epidemic situation outside Hubei Province and online rumors. The peak correlation was 0.59 ($P<.001$) when the lag value was 0 days, indicating that the daily number of newly confirmed cases was significantly positively correlated with rumors, with changes occurring simultaneously. Figure 2B shows the relationship between the COVID-19 epidemic situation in Hubei Province and online rumors. The peak correlation was 0.55 ($P<.001$) with a lag of -1 day, which suggests that the total number of rumors changed 1 day in advance of the change in the number of cases in Hubei Province. In addition, the epidemic situation within and outside Hubei Province was not significantly related to the 4 types of emotional rumors.

Figure 2. (A) Cross-correlation between the daily number of newly confirmed cases outside Hubei Province and all rumors. (B) Cross-correlation between the daily number of newly confirmed cases in Hubei Province and all rumors. Blue dashed lines denote the 95% confidence intervals of the uncorrelated time series. HB: Hubei Province.
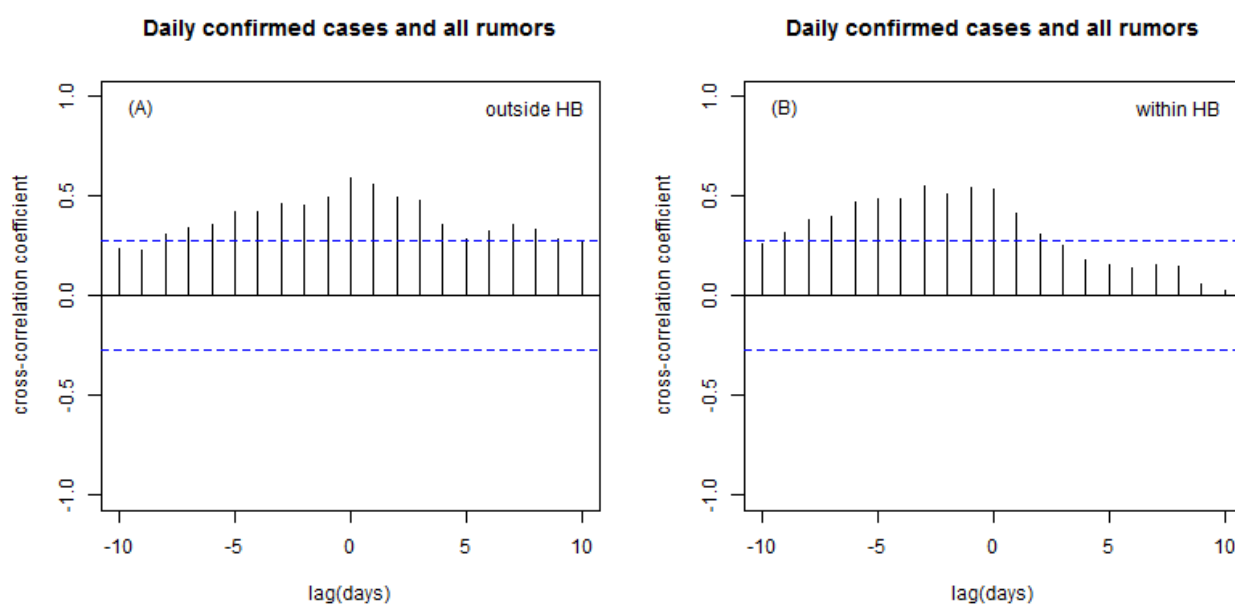

Relationship Between Public Emotions and Rumors

During the study period, anger (2,248,362/17,328,675, 12.97%) dominated public emotions on the internet, followed by fear (627,407/17,328,675, 3.62%), happiness (216,072/17,328,675, 1.25%), and sadness (195,708/17,328,675, 1.13%). Figure 3 presents the trend of public emotions on Weibo, which shows that the number of anger-related and fear-related comments were more than the number of happiness-related and sadness-related comments almost every day.

Figure 3. Trend of public emotions expressed on Weibo in the context of COVID-19 during the study period.
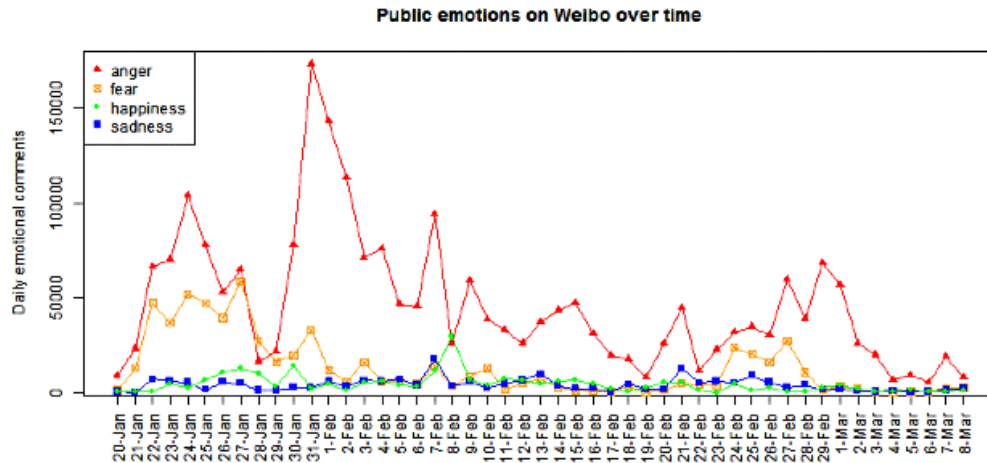
Next, we found that the peak correlation of anger and rumors was 0.48 (*P*<.001) with a lag of 0 day (Figure 4A). That is, the angrier the public got, the more rumors there would be. Similarly, the time-lagged analysis between fear and rumors (Figure 4B) and between sadness and rumors (Figure 4D) also indicated significant positive correlations among these 2

emotions and rumors. The results showed the correlations for both these emotions reached their peaks of 0.51 (*P*<.001) and 0.47 (*P*<.001), respectively, with the lag of 0 days. Additionally, a positive correlation was observed between happiness and rumors (r=0.56, *P*<.001, Figure 4C), with a time lag of 1 day. That is, happiness would lag the emergence of rumors by 1 day.

Figure 4. Cross-correlation between emotions and all rumors. Cross-correlation between (A) anger and all rumors, (B) fear and all rumors, (C) happiness and all rumors, and (D) sadness and all rumors. Blue dashed lines denote the 95% confidence intervals of the uncorrelated time series.

XSL•FO
RenderX
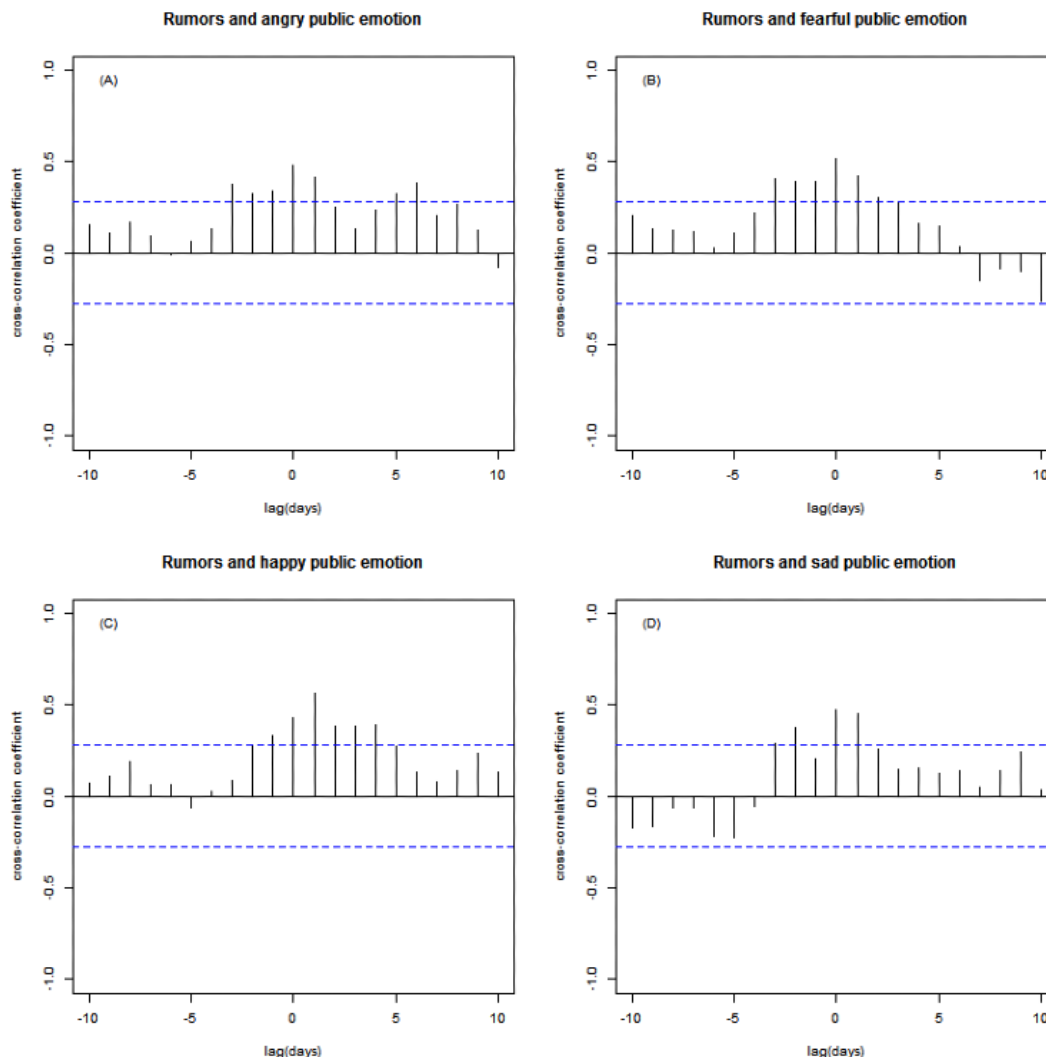
Finally, we analyzed the relationship between the 4 types of emotions and different types of rumors. Figure 5 shows that there was a positive correlation only between fear and fearful

rumors, with the peak correlation (r=0.34, *P*=.02)) observed at the lag of 0 day.

Figure 5. Cross-correlation between fear and fearful rumors. Blue dashed lines denote the 95% confidence intervals of the uncorrelated time series.
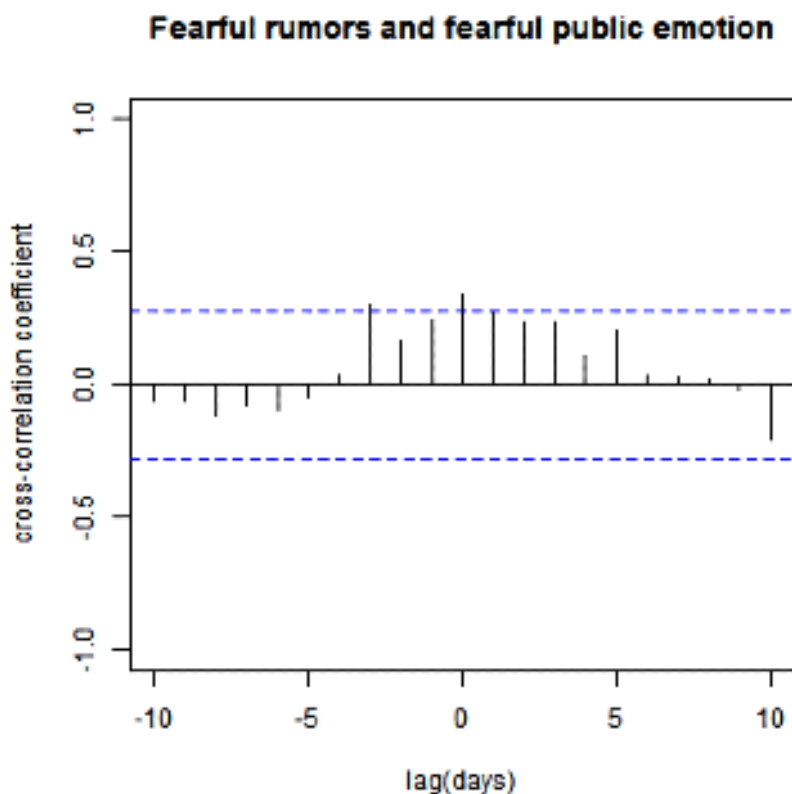
Discussion

Principal Results

To our knowledge, this is the first study to explore the relationship between public emotions and rumors during COVID-19 by using a web-based monitoring method. A large number of comments and rumors related to COVID-19 were obtained from the Weibo account of People's Daily and Tencent Myth Busters by using the web crawler Scrapy. This web-based monitoring study overcomes the limitations of traditional survey methods and facilitates data-collection in a rapid and real-time manner during outbreak of a public health emergency. In particular, during a large-scale quarantine period, people are more likely to use the internet and social media to acquire and propagate the latest information about the epidemic [7], as well as actively participate in related topic discussions [27]. The internet and social media have become important platforms for real-time monitoring and assessment of changes in public emotions and behaviors. In addition, compared with traditional methods such as survey and telephone interviews, web-based monitoring tends to be more objective and authentic, effectively avoiding measurement errors caused by self-reporting or recalling [14,28]. Manual identification was employed to code emotional categories, which also improved the classification accuracy of public emotion identification [22,23].

In this study, statistically significant relationships were first observed among the daily number of newly confirmed cases and the total number of rumors both within and outside of Hubei

Province, which is in line with a previous study that found that the uncertainty of sudden infectious diseases can lead to a large number of rumors [18]. In another study, Li et al [29] analyzed the emotional reactions of 4607 participants through an online survey and found that the COVID-19 did not lead to changes in the frequency of public emotions. However, our study found that the public expressed more anger and fear than happiness and sadness, and there were several large fluctuations in public emotions. Some researchers have suggested that rumors are an outlet for the public to express fear or pressure after a disaster, and spreading rumors could become an effective way for the public to express their emotions such as anger, fear, and sadness [30]. Our findings are in line with these previous findings. The results of our study revealed that three emotions (ie, anger, fear, and sadness) were positively correlated to the total number of rumors, and the emotions and rumors changed synchronously. Moreover, we found that the increase in the happiness emotion tended to lag the emergence of rumors by 1 day. Previous studies have found that public emotions usually transform from negative to positive after the rumors are disproved [31]. In line with these findings, we observed that when the rumors were verified and clarified, there was a proportional increase in the public's happiness.

Regarding the relationship between the 4 basic human emotions and 4 different types of emotional rumors, our results showed that fearful rumors were positively associated with fear. This result supports the argument presented by of Na et al [8], which claims the consistency between an individual's emotional state and the emotional information in rumors could increase the

prevalence of the rumors. However, our findings regarding anger and anger-related rumors were inconsistent with their conclusion that anger could lead the public to accept anger-related rumors [8]. This difference might be attributed to the differences of sample choice: our study focused on a Chinese sample, whereas Na et al [8] analyzed an American sample; hence, cultural differences in the relationship between public emotions and rumors also need to be further studied.

Study Limitations

This study, however, has several limitations. First, this study only focused on Sina Weibo for data collection. Other media platforms, such as WeChat in China and Twitter in other countries, were not included to acquire more extensive data. Second, batch-fetching data can only be obtained as a whole; hence, we cannot investigate the change in public emotions at an individual level. In future research, we may use other methods such as surveys or interviews to examine the trajectories of public emotions at an individual level. Third, this study analyzed the relationship between rumors and 4 basic

human emotions, and further research could be extended to examine the correlations between rumors and more complex emotions such as anxiety.

Conclusions

Our findings provide several suggestions for relevant authorities and policy makers in guiding emotions of the public during public health emergencies. First, during a large-scale quarantine period, the authorities can use web-based monitoring methods to identify public emotions and behaviors in real time and provide timely guidance to channel public emotions and behaviors. Second, rumors are a catalyst for public emotions, and disproving them in a timely manner would be helpful to increase positive emotions of the public. Third, our findings showed that fearful rumors were associated with fear. Thus, media platforms should strengthen the monitoring of online rumors, identify and verify emotional rumors in a timely manner, and minimize the spread of fearful rumors to reduce fear among the public.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus (2019-nCoV) Situation Report - 10. 2020. World Health Organzation. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2 [accessed 2020-04-28]
2. Xiao C. A novel approach of consultation on 2019 novel coronavirus (COVID-19)-related psychological and mental problems: structured letter therapy. Psychiatry investigation 2020;17(2):175. [doi: 10.30773/pi.2020.0047] [Medline: 32093461]
3. Rosnow R, Foster E. Rumor and gossip research. Psychological Science Agenda 2005;19(4):1-2.
4. Prasad J. The psychology of rumour: a study relating to the great Indian earthquake of 1934. British Journal of Psychology 1935 Jul;26(1):1-15 [FREE Full text] [doi: 10.1111/j.2044-8295.1935.tb00770.x]
5. Allport GW, Postman L. The Psychology of Rumor. New York: Henry Holt and Company; 1947.
6. Marett K, Joshi K. The decision to share information and rumors: examining the Role of motivation in an online discussion forum. CAIS 2009 Feb;24(1):4. [doi: 10.17705/1cais.02404]
7. Chua A, Tee C, Pang A, Lim EP. The Retransmission of Rumor-related Tweets: Characteristics of Source and Message. 2020 Jul Presented at: SMSociety '16: Proceedings of the 7th 2016 International Conference on Social Media & Society; July 11-13, 2016; London, United Kingdom p. 1-10. [doi: 10.1145/2930971.2930994]
8. Na K, Garrett RK, Slater MD. Rumor acceptance during public health crises: testing the emotional congruence hypothesis. Journal of Health Communication 2018 Oct 16;23(8):791-799. [doi: 10.1080/10810730.2018.1527877]
9. Tasnim S, Hossain M, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
10. Yang JZ, Chu H. Who is afraid of the Ebola outbreak? The influence of discrete emotions on risk perception. Journal of Risk Research 2016 Oct 28;21(7):834-853. [doi: 10.1080/13669877.2016.1247378]
11. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Inectious Diseases/SARS Community Outreach Team. Fear and stigma: the epidemic within the SARS outbreak. Emerg Infect Dis 2004 Feb;10(2):358-363 [FREE Full text] [doi: 10.3201/eid1002.030750] [Medline: 15030713]
12. Qian M, Dongmei YE, Dong W. Behaviour, cognition and emotion of the public in Beijing towards SARS. Chinese Mental Health Journal 2003;17(8):515-520. [doi: 10.1023/A:1022289509702]
13. Liao Q, Wu P, Wing Tak Lam W, Cowling BJ, Fielding R. Trajectories of public psycho-behavioural responses relating to influenza A(H7N9) over the winter of 2014-15 in Hong Kong. Psychol Health 2019 Feb;34(2):162-180. [doi: 10.1080/08870446.2018.1515436] [Medline: 30430862]

14. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys. BMC Public Health 2011 Jan 03;11:2 [FREE Full text] [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]

15. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. PLoS Negl Trop Dis 2017 Mar;11(3):e0005354 [FREE Full text] [doi: 10.1371/journal.pntd.0005354] [Medline: 28263988]

16. Liu K, Wang T, Yang Z, Huang X, Milinovich GJ, Lu Y, et al. Using Baidu Search Index to Predict Dengue Outbreak in China. Sci Rep 2016 Dec 01;6:38040 [FREE Full text] [doi: 10.1038/srep38040] [Medline: 27905501]

17. The 45th China Statistical Report on Internet Development in China in Chinese. China Internet Network Information Center. 2020. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202004/P020200428399188064169.pdf [accessed 2020-04-28]

18. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]

19. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]

20. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

21. Festinger L. A Theory of Cognitive Dissonance. Vol. 2. Redwood City, California, USA: Stanford University Press; 1957.

22. Singh NK, Tomar DS, Sangaiah AK. Sentiment analysis: a review and comparative analysis over social media. J Ambient Intell Human Comput 2018 May 23;11(1):97-117. [doi: 10.1007/s12652-018-0862-8]

23. Mäntylä MV, Graziotin D, Kuutila M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review 2018 Feb;27:16-32. [doi: 10.1016/j.cosrev.2017.10.002]

24. Jack RE, Sun W, Delis I, Garrod OGB, Schyns PG. Four not six: revealing culturally common facial expressions of emotion. J Exp Psychol Gen 2016 Jun;145(6):708-730. [doi: 10.1037/xge0000162] [Medline: 27077757]

25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]

26. Tencent Jiaozhen platform. 2020. URL: https://vp.fact.qq.com/home [accessed 2020-04-28]

27. Henrich N, Holmes B. What the public was saying about the H1N1 vaccine: perceptions and issues discussed in on-line comments during the 2009 H1N1 pandemic. PLoS One 2011 Apr 18;6(4):e18479 [FREE Full text] [doi: 10.1371/journal.pone.0018479] [Medline: 21533161]

28. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]

29. Li J, Yang A, Dou K, Wang L, Zhang M, Lin X. Chinese public's knowledge, perceived severity, and perceived controllability of COVID-19 and their associations with emotional and behavioural reactions, social participation, and precautionary behaviour: a national survey. BMC Public Health 2020 Oct 21;20(1):1589 [FREE Full text] [doi: 10.1186/s12889-020-09695-1] [Medline: 33087109]

30. Rosnow RL. Rumor as communication: a contextualist approach. Journal of Communication 1988 Mar;38(1):12-28. [doi: 10.1111/j.1460-2466.1988.tb02033.x]

31. Zeng R, Zhu D. A model and simulation of the emotional contagion of netizens in the process of rumor refutation. Sci Rep 2019 Oct 02;9(1):14164 [FREE Full text] [doi: 10.1038/s41598-019-50770-4] [Medline: 31578405]

Abbreviations

NHC: The National Health Commission of People's Republic of China

XSL•FO
RenderX

Original Paper

The Public's Perception of the Severity and Global Impact at the Start of the SARS-CoV-2 Pandemic: A Crowdsourcing-Based Cross-Sectional Analysis

Orr Shauly[1*], BSc; Gregory Stone[1*], BA; Daniel Gould[1*], MD, PhD

Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[*]all authors contributed equally

Corresponding Author:
Daniel Gould, MD, PhD
Keck School of Medicine
University of Southern California
1450 San Pablo Street
Suite 415
Los Angeles, CA
United States
Phone: 1 323 442 7920
Email: dr.danjgould@gmail.com

Abstract

Background: COVID-19 is a rapidly developing threat to most people in the United States and abroad. The behaviors of the public are important to understand, as they may have a tremendous impact on the course of this novel coronavirus pandemic.

Objective: This study intends to assess the US population's perception and knowledge of the virus as a threat and the behaviors of the general population in response.

Methods: A prospective cross-sectional study was conducted with random volunteers recruited through Amazon Mechanical Turk, an internet crowdsourcing service, on March 24, 2020.

Results: A total of 969 participants met the inclusion criteria. It was found that the perceived severity of the COVID-19 pandemic significantly differed between age groups ($P<.001$) and men and women ($P<.001$). A majority of study participants were actively adhering to the Centers for Disease Control and Prevention guidelines.

Conclusions: Though many participants identified COVID-19 as a threat, many failed to place themselves appropriately in the correct categories with respect to risk. This may indicate a need for additional public education for appropriately defining the risk of this novel pandemic.

Introduction

A novel coronavirus, COVID-19, has resulted in an ongoing global pandemic marked by a viral pneumonia with severe morbidity and mortality in 3%-5% of those infected [1,2]. With cases mounting in the United States over the past month and with the Centers for Disease Control and Prevention (CDC) and many local governments issuing safety measures including social distancing and cleanliness guidelines, the public has had to adapt to a new way of life [3,4]. However, these drastic changes may have a tremendous psychosocial impact.

The public's perception of this pandemic's severity may also impact adherence to CDC guidelines or regionally mandated quarantine, and suboptimal adherence to guidelines may have a detrimental effect on attempts to curb the continued spread of the COVID-19 pandemic. It is well known across many chronic health conditions (asthma or cardiovascular disease) and screening health tools such as colonoscopy and mammography that patient's perception of severity or risk may directly correlate with adherence to treatment and guidelines [5-9]. It is thus important to survey the public's perception of the pandemic's

severity and its psychosocial impact across age, sex, and individual risk factors.

Methods

Cross-Sectional Study

A prospective cross-sectional study of random volunteers recruited through an internet crowdsourcing service, Amazon Mechanical Turk (MTurk), was conducted. Study participants were recruited and performed the survey on March 24, 2020. Individuals were required to be older than 18 years and registered through the Amazon service platform within the United States to prevent workers from taking the same survey multiple times or from outside the United States, and the survey was not advertised. Internet crowdsourcing is a powerful tool in its ability to quantify the perceptions and opinions of a diverse group of individuals that may otherwise be inaccessible through other surveying methodologies at a fraction of the cost [10-13]. Several studies have shown that the participant population is extremely representative of the United States internet population, with 70%-80% of all workers originating and recording responses from the continental United States [11,14]. In this study, 100% of the workers were from the United States, as an Internet Protocol filter and registration filter were used to guarantee US response. Workers were provided with a compensation of US $0.25 per response (limit one per worker) and were screened by the platform for attentiveness to detail and completeness. Using the MTurk platform to administer surveys significantly reduces cost, response time, inaccuracy, and barriers to access for specific patient populations, providing a more holistic snapshot of the overall US population.

Responses were crowdsourced for a 5-minute open survey that collected data on several themes: opinions on travel, CDC guidelines and adherence, ethical concerns, resource use, proactive actions, social distancing, and public perception of several facets of the COVID-19 pandemic in the United States (Multimedia Appendix 1). Regarding the public perception of severity, concern, and financial impact, study participants were asked to provide a utility score on a 10 cm vertical visual analog scale (VAS) between 0 cm-10 cm, with the scale described in detail for each question to maximize clarity of the scale being used. All collected utility scores were analyzed across age, sex, household income, and individual health risk factors such as the presence of chronic comorbidities. Questionnaire items were not randomized, and adaptive questions were used to reduce the number and complexity of survey questions. There were 2-12 questionnaire items per page, and the survey was distributed over 11 pages. Participants were able to review and change their answers. Only completed surveys were analyzed, and completeness was assessed after all surveys were collected.

Institutional review board approval was not sought or necessary for this study, as in the case of MTurk, no identifiable private or personal information was recorded and no direct interaction or engagement with any of the study participants occurred. The methodological soundness of the manuscript was assessed using the CHERRIES (Checklist for Reporting Results of Internet E-Surveys; Multimedia Appendix 2) [15].

Screening Questions

Although Amazon MTurk requires that registered volunteers be older than 18 years, individuals may not be providing accurate demographic information when creating their account. To ensure that all surveyed participants were considered adults, the first questions of the survey asked the participants to confirm their age. No other screening questions were administered to maintain a truly diverse representation of the US population.

Attention Check Question

To ensure that survey participants were paying close attention to each question and scenario, and that the generated data was a valid representation of patient opinions, the following attention check mechanism was included at two points in the survey [16]. Study participants were asked in two different sections (on separate pages, where participants could not go backwards to change their answer) whether they were currently taking Plaquenil. Any discordance between the two questions resulted in exclusion from this study. Those that were excluded were prevented from ever taking this survey again.

Statistics

Data from the survey was collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) and pooled and analyzed using Microsoft Excel 2016 (Microsoft Corporation). Statistics were performed using R Statistical Package (R Foundation for Statistical Computing) with continuous data evaluated using two-tailed two-sample unequal variances t tests (α=.05) assuming a normal distribution for comparisons between two groups and with analysis of variance with post hoc analysis with Tukey test, which allows for pairwise comparisons for comparisons between more than two groups. Categorical data was evaluated using chi-square testing. All responses were also screened by the Amazon MTurk platform and manually by the authors using the REDCap platform for completeness, validity, and duplicate responses.

Results

Study Participants

A total of 1107 MTurk participants were surveyed. Of these, 64 (5.8%) were excluded because they did not complete the survey, and 40 (3.6%) were excluded because they failed the attention check question. An additional 17 individuals were screened to have entered duplicate responses (identified by worker ID on the MTurk platform) and were deleted from the REDCap database (34 total responses). A total of 969 participants met inclusion criteria, were screened for validity and completeness, did not fail the attention check question, and were included in this study.

The MTurk human intelligence platform aims to capture a diverse number of study participants across age, sex, ethnicity, and income. Our study participants included 410 males and 559 females. The majority of the 969 participants were between the ages of 18-39 years (n=538, 55.5%), with the remaining between the ages of 40-64 years (n=366, 37.8%) or older than 65 years (n=65, 6.7%). The survey also captured a diverse representation of the United States—participants were from every state except

for Montana and Vermont, with the majority from California, Florida, Texas, and New York (Figure 1).

Figure 1. Distribution of survey respondents across the United States, with color representation of total survey impact: dark green represents the largest number of survey participants (California, n=100) and light beige the fewest with only a single respondent (Wyoming and North Dakota). There were no respondents from Montana or Vermont (white).
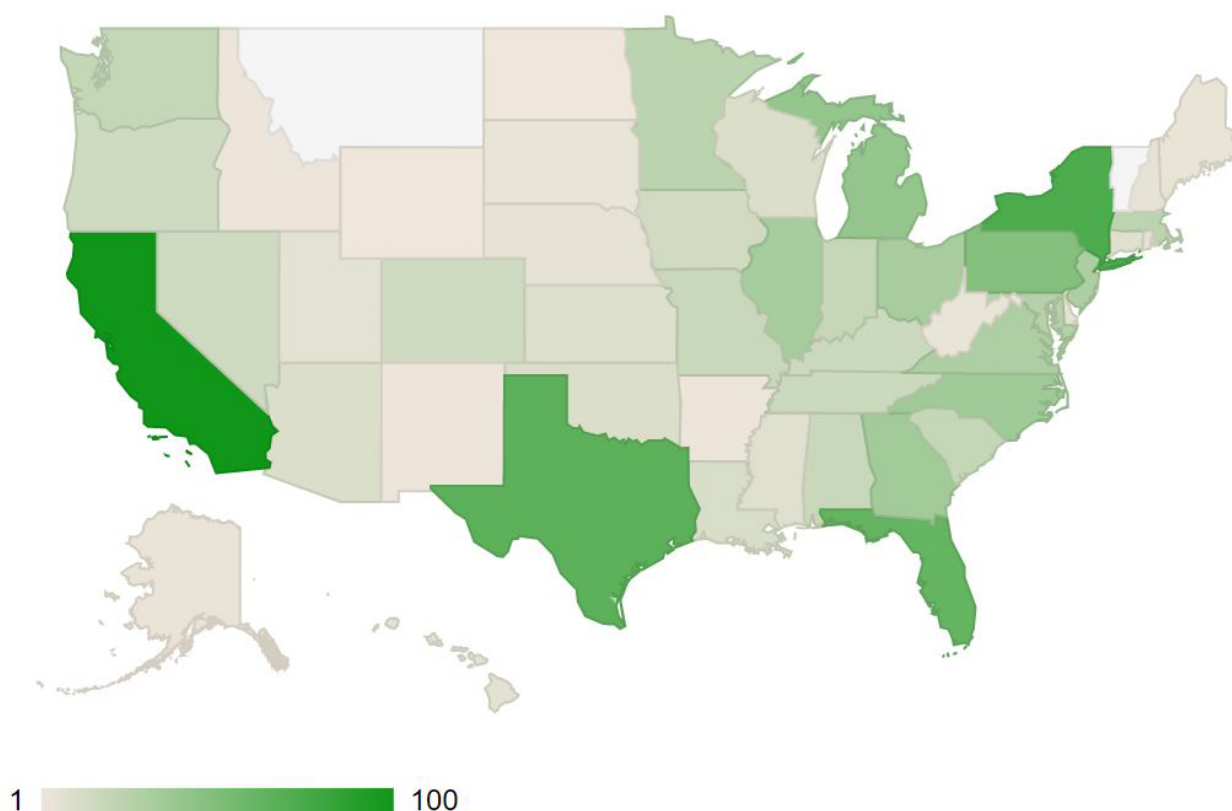
Ethnicity was derived using the guidelines set forth by the US census bureau. Of the 969 participants, a total of 11 (1.1%) participants identified as American Indian or Alaskan Native, 91 (9.4%) as Asian, 108 (11.1%) as Black or African American, and 753 as White or Hispanic (77.6%). With respect to ethnicity, our subject population is extremely representative of the US population, as the 2019 US census demonstrated 76.3% of the population as White or Hispanic, 13.4% as Black or African American, 5.9% as Asian, and 1.3% as American Indian or Alaska Native (Multimedia Appendix 3) [17].

In addition to both accurate geographic and cultural representation with respect to the US population, the study demonstrated a median household income of US $52,000 across all study participants, which is similar to the 2019 census bureau figure of US $60,293 [17]. The distribution of income across all study participants can be viewed in Multimedia Appendix 4.

Study participants were asked to disclose pertinent medical history that may put them at higher risk of morbidity and mortality from COVID-19. Out of 969 participants, a total of 367 (37.9%) indicated they had a chronic medical condition including diabetes, high blood pressure, high cholesterol, chronic obstructive pulmonary disease (COPD), asthma, HIV/AIDS, or chronic heart disease. An additional 145 (15.0%) participants indicated a history of one or more autoimmune conditions including psoriasis, lupus, rheumatoid arthritis, Crohn disease, multiple sclerosis, alopecia areata, sarcoidosis, myasthenia gravis, systemic sclerosis, pemphigus, ankylosing spondylitis, or other conditions not listed (Figure 2). Participants did not divulge whether the disease was currently active. However, 49 (5.1%) indicated they were currently taking oral steroid medication, 22 (2.27%) participants indicated they were a recipient of a recent organ transplantation requiring current immunosuppressive medication, and 18 (1.86%) subjects indicated they were currently being treated with chemotherapy. There were 187 (19.3%) participants self-identified as currently smoking tobacco. In total, 508 (52.4%) high-risk individuals were identified.

Figure 2. Total number of survey participants that indicated they have either a chronic health condition (n=390) or an autoimmune disorder that may increase susceptibility and risk to COVID-19 infection (n=169), and subsequent morbidity and mortality, in comparison to the ten most common chronic health conditions among persons living in residential care facilities (as of the national survey of residential care facilities in the United States in 2010), published by the Centers for Disease Control and Prevention on August 10, 2012, as reported by Caffrey et al (upper right). Darker red corresponds to the most common and lighter red to the least common chronic health conditions. COPD: chronic obstructive pulmonary disease.
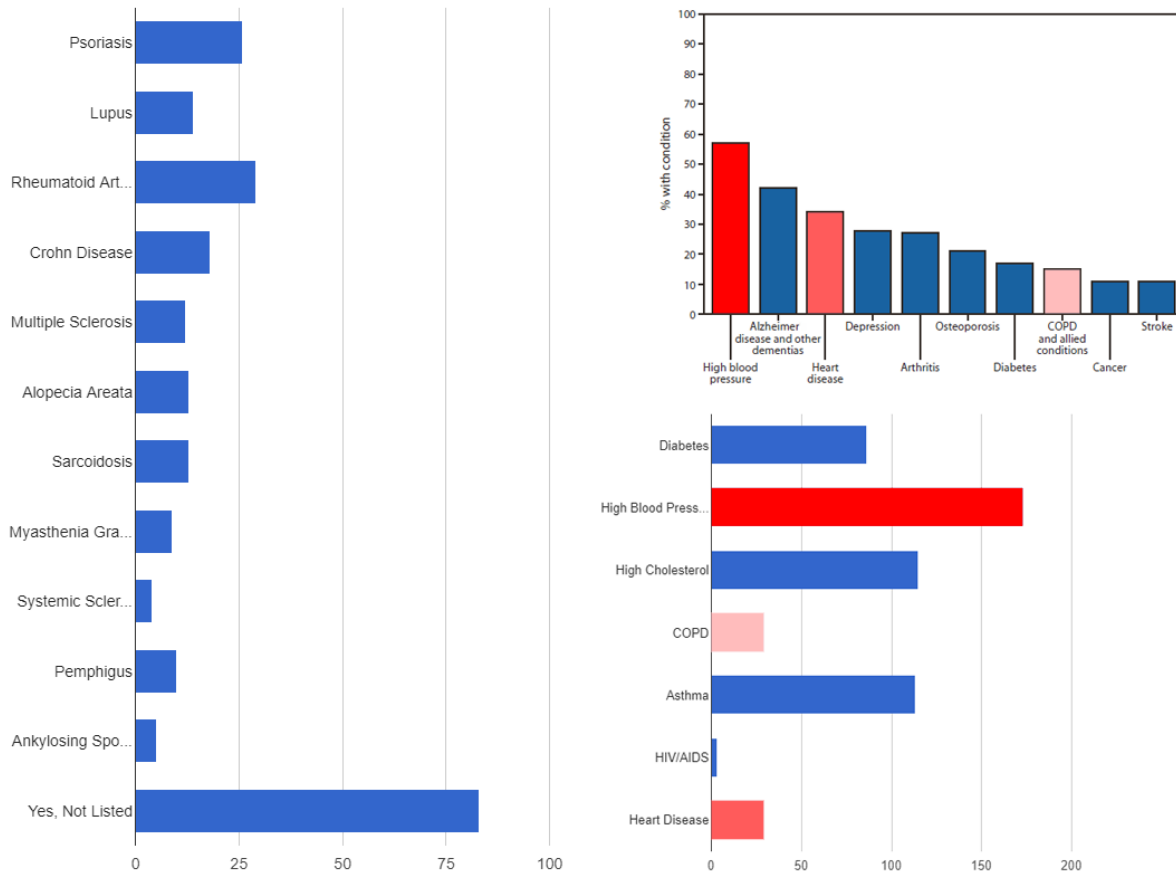
Mental Health Impact

Study participants were asked if social distancing has had a negative impact on their mental health (Figure 3). Pearson chi-square test demonstrated a significant mental health impact in those aged 18-39 years when compared to those ≥40 years ($P<.001$). Pearson chi-square test with Yates continuity correction also demonstrated a significantly higher negative impact on the mental health of female participants as compared to males ($P=.01$).

Figure 3. Survey response to "has social distancing had a negative impact on your mental health" question across sex and age cohorts, and a scatter plot depicting the visual analog scale scores of how effective people believe social distancing is at addressing the spread of COVID-19 (where 0 is not effective at all, and 10 is most effective), with a mean of 7.98 (SD 1.94) and a median of 8.40 (red dot).
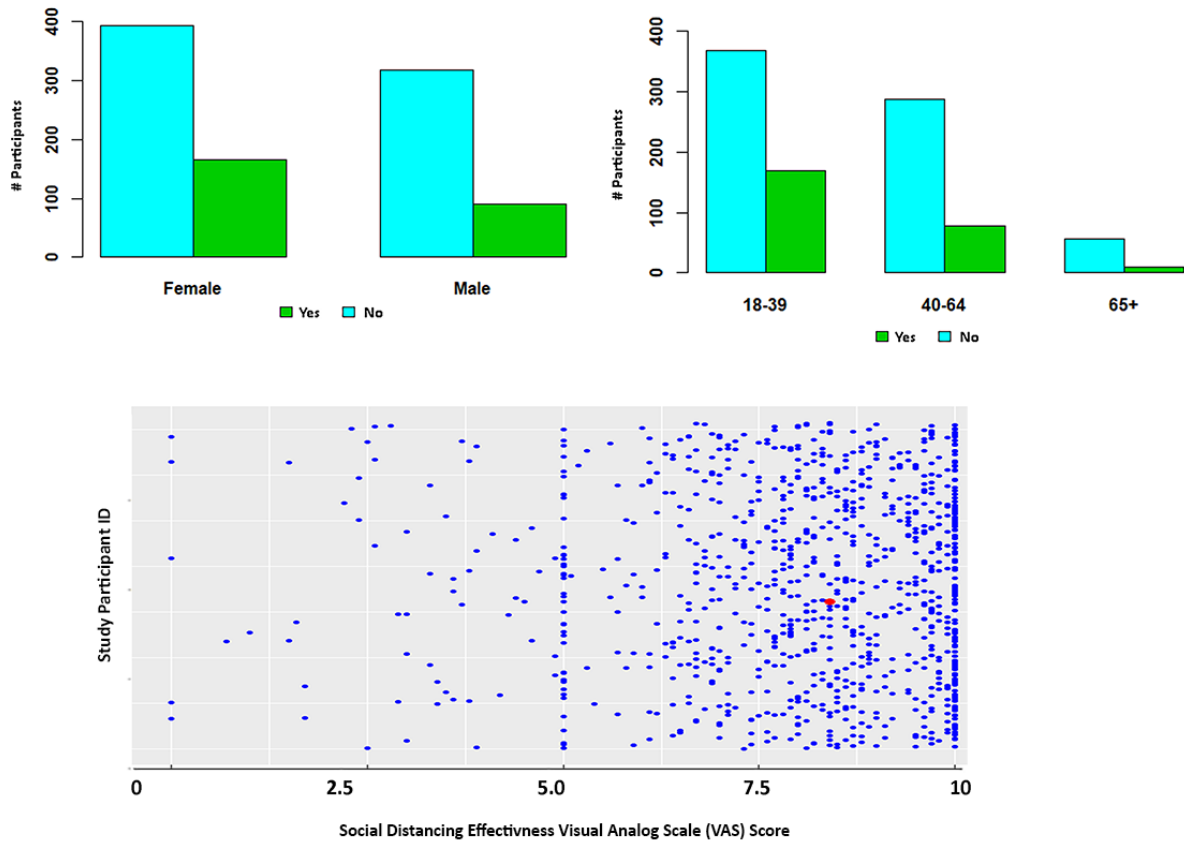
Severity Perception

Study participants were asked to indicate on a VAS scale (from 0 to 10, where 10 represented the highest level of severity) how severe they believe the COVID-19 outbreak to be, considering published data from the CDC and their own perceptions of the severity (Figure 4). Participants were provided an infographic that demonstrated the number of cases across the United States and information about the case mortality rate.

Figure 4. Box and whisker plots depicting the distribution and median of visual analog scale scores across all study participants with respect to age cohorts in response to how concerned individuals are about the negative impact of COVID-19 on the overall global economy (with 0 representing no concern at all and 10 the most concern), how concerned individuals are about the negative impact of COVID-19 on the overall US economy, how concerned individuals are about the negative impact of COVID-19 on personal financial stability, and how appropriately do individuals believe the US government has responded to the COVID-19 pandemic (where 0 indicates that they do not think the government or Centers for Disease Control and Prevention [CDC] has responded appropriately at all and 10 indicates that they think the government and CDC have responded perfectly).
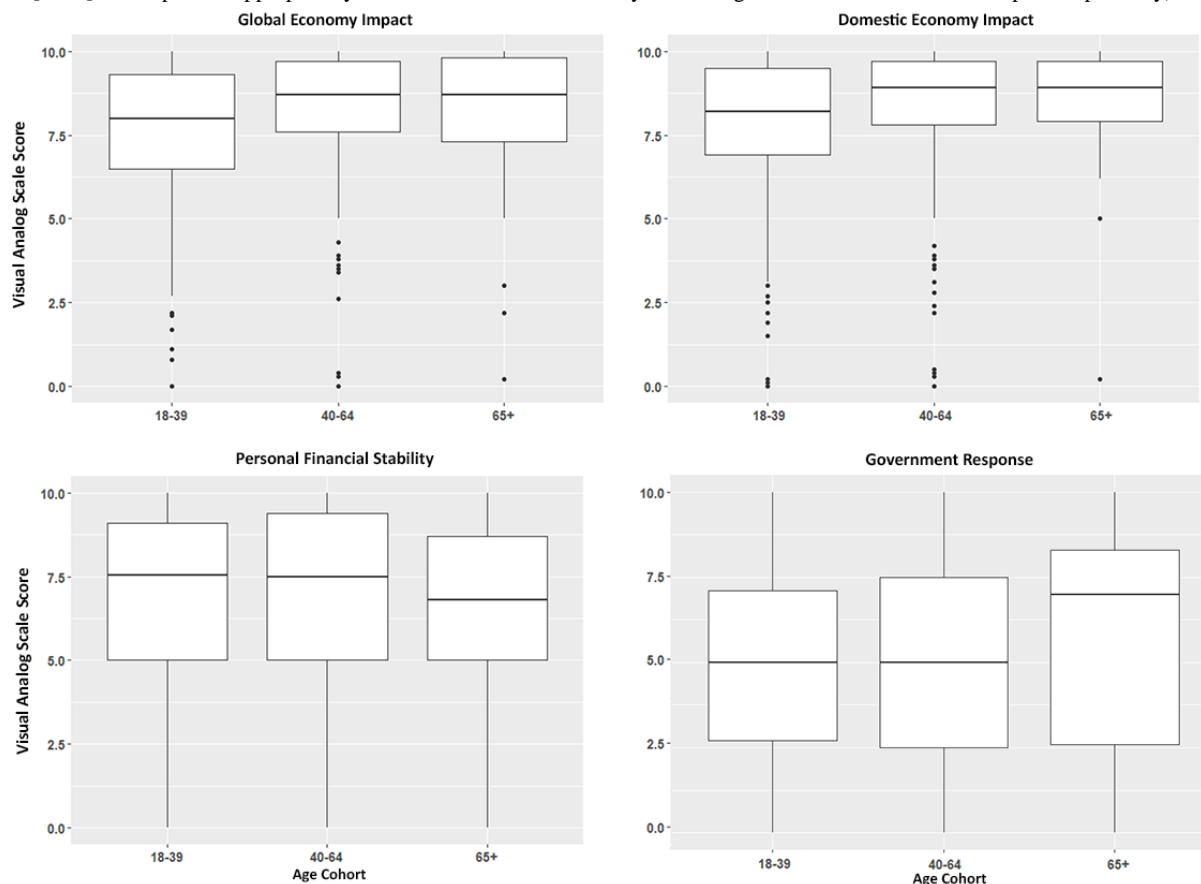
It was found that age had a linear correlation to perceived severity (F_2=8.21, P<.001). Those aged 18-39 years indicated a mean VAS of 7.05 (95% CI 6.88-7.22, the lowest severity), those aged ≥65 years reported a mean VAS of 8.06 (95% CI 7.59-8.54, the highest severity), and those aged 40-64 years demonstrated a mean VAS of 7.40 (95% CI 7.17-7.63). The difference in perceived severity was statically significant when comparing those younger than 40 years to those 65 years or older (P<.001) and when comparing those younger than 40 years to those aged 40-64 years (P=.04). The difference in perceived severity did not differ between those aged 40-65 years and those aged ≥65 years (P=.05). Reported worry about personal health (indicated by rating on VAS scale from 0 to 10, where 0 represented no worry and 10 represented extreme worry) did not significantly differ across age groups (P=.13). Those aged 18-39 years indicated a mean VAS of 5.73 (95% CI 5.50-5.96), those aged 40-64 years reported a mean VAS of 6.07 (95% CI 5.78-6.36), and those aged ≥65 years demonstrated a mean VAS of 6.17 (95% CI 5.45-6.90).

It was found that the perceived severity of the COVID-19 pandemic significantly differed between men and women ($t_{834.22}$=3.942, P<.001). Men indicated a mean VAS of 6.94 (95% CI 6.72-7.15), and women indicated a mean VAS of 7.48 (95% CI 7.31-7.65).

The perceived severity of the COVID-19 pandemic did not differ between high-risk and low-risk participants ($t_{966.91}$=−1.1766, P=.24), where high risk was defined as a participant with any of the following: at least one chronic medical condition including diabetes, high blood pressure, high cholesterol, COPD, asthma, HIV/AIDS, or chronic heart disease; a history of any autoimmune condition including psoriasis, lupus, rheumatoid arthritis, Crohn disease, multiple sclerosis, alopecia areata, sarcoidosis, myasthenia gravis, systemic sclerosis, pemphigus, ankylosing spondylitis, or other conditions not listed; current oral steroid medication use; recipient of a recent organ transplantation requiring current immunosuppressive medication; current treatment with chemotherapy; or current tobacco smoker. High-risk participants indicated a mean VAS of 7.33 (95% CI 7.13-7.52) and low-risk participants reported a mean VAS of 7.17 (95% CI 6.98-7.35) with respect to perceived severity. However, high-risk individuals were more likely to believe that they would become infected with COVID-19 (P=.005) and experience serious illness or die due to infection with COVID-19 (P<.001). Furthermore, 32.7% (166/508) of high-risk individuals and 24.3% (112/461) of low-risk individuals believed they would become infected with COVID-19, and 15.6% (79/508) of high-risk individuals and 7.81% (36/461) of low-risk individuals believed they would become seriously ill or die due to COVID-19.

When stratifying participants by annual income (US $0-US $25,000; US $25,001-US $50,000; US $50,001-US $75,000; US $75,001-US $100,000; and greater than US $100,000), no significant difference in perceived severity was observed across any income cohort (F_4=0.33, P=.86).

Government Response Perception

Study participants were asked to provide their opinion regarding the appropriateness of the US government's response to the novel coronavirus pandemic (as of March 24, 2020) by way of a VAS scale from 0 to 10, where 0 represented an inappropriate response and 10 represented an appropriate response. It was found that the perceived appropriateness of the US government's response significantly differed across age groups (F_2=3.663, P=.03). Those aged 18-39 years indicated a mean VAS of 4.85 (95% CI 4.62-5.08, the lowest appropriateness), those age ≥65

years reported a mean VAS of 5.78 (95% CI 4.96-6.60, the highest appropriateness), and those aged 40-64 years demonstrated a mean VAS of 5.16 (95% CI 4.86-5.46). The difference in perceived appropriateness of government response was statically significant when comparing those younger than 40 years to those aged ≥65 years (P=.04) but not when comparing those younger than 40 years to those aged 40-64 years (P=.25) or those aged 40-64 years to those aged ≥65 years (P=.24).

Economic Impact

Study participants were asked to indicate on a VAS scale (from 0 to 10, where 0 represented no concern and 10 represented the highest level of concern) their concern regarding the effect of the COVID-19 outbreak on their personal financial stability, the US economy, and the global economy (Figure 5).

Figure 5. Box and whisker plots depicting the distribution and median of visual analog scale scores of study participants in response to their perception of the severity of the COVID-19 pandemic, with respect to age cohort, sex, income level, and health risk status, with high-risk individuals defined within the text.
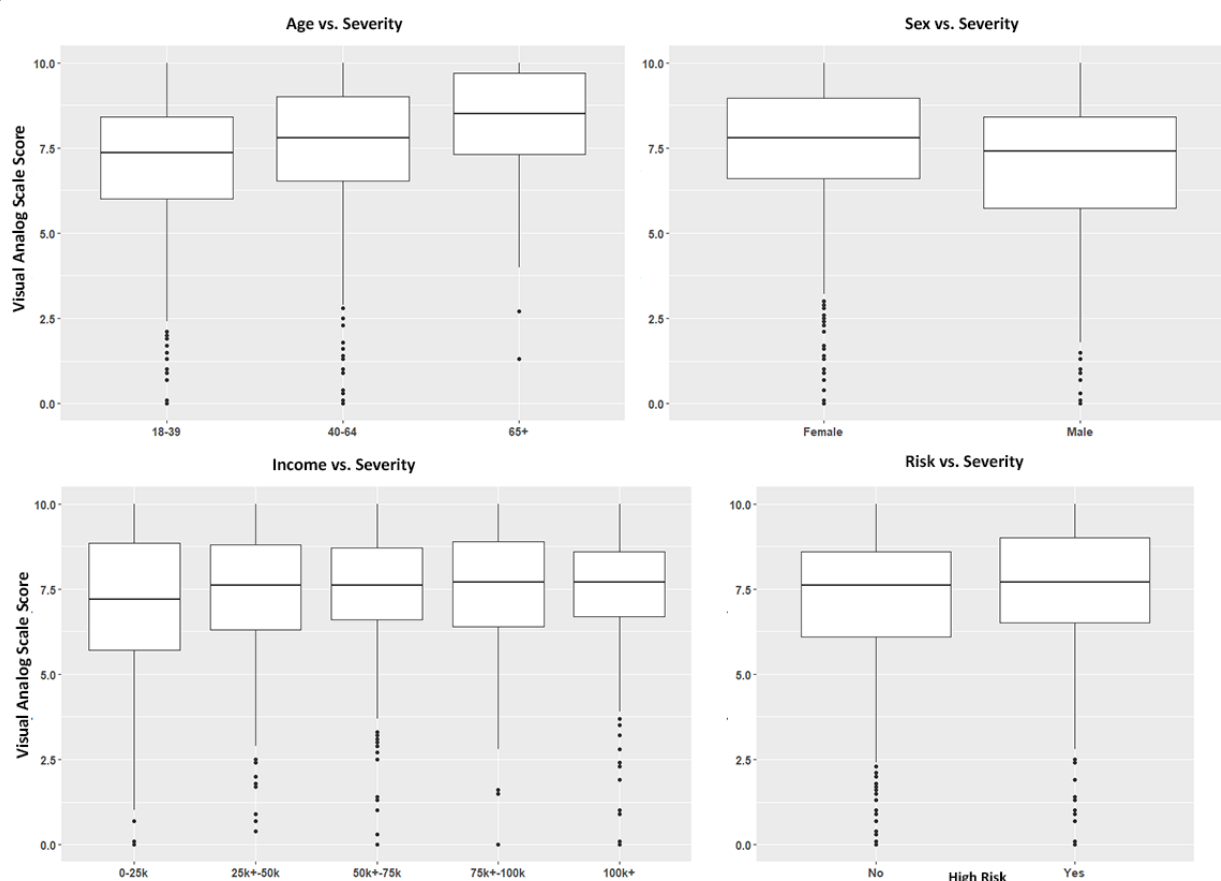
The reported concern regarding personal financial stability was not found to differ significantly between age groups (F_2=0.743, P=.48). Those aged 18-39 years reported a mean VAS of 6.96 (95% CI 6.74-7.18), those aged 40-64 years indicated a mean VAS of 6.92 (95% CI 6.64-7.20), and those aged ≥65 years demonstrated a mean VAS of 6.53 (95% CI 5.85-7.21).

The reported concern with respect to the effect of COVID-19 on the US economy was found to significantly differ between age groups (F_2=11.74, P<.001). Those aged 18-39 years indicated a mean VAS of 7.86 (95% CI 7.69-8.03, lowest concern), those aged ≥65 years indicated a mean VAS of 8.52

(95% CI 8.12-8.93, highest concern), and those aged 40-64 years indicated a mean VAS of 8.43 (95% CI 8.25-8.61). The difference in perceived concern regarding the US economy was statically significant when comparing those younger than 40 years to those aged 40-64 years (P<.001) and when comparing those younger than 40 years to those aged ≥65 years (P=.02). When comparing those aged 40-64 years to those aged ≥65 years, no significant difference was observed (P=.93).

The reported concern regarding the effect of COVID-19 on the global economy was found to significantly differ between age groups (F_2=14.61, P<.001). Those aged 18-39 years indicated

a mean VAS of 7.64 (95% CI 7.47-7.81, lowest concern), those aged 40-64 years reported a mean VAS of 8.33 (95% CI 8.14-8.51, highest concern), and those aged ≥65 years demonstrated a mean VAS of 8.23 (95% CI 7.75-8.71). The difference in perceived concern regarding the global economy was statically significant when comparing those younger than 40 years to those aged 40-64 years (*P*<.001). When comparing those younger than 40 years to those aged ≥65 years, or when comparing those aged 40-64 years to those aged ≥65 years, no significant difference was observed (*P*=.05 and *P*=.92, respectively).

Social Distancing

Social distancing from family, friends, and coworkers was not observed to be significantly associated with age groups (*P*=.15). Of the 538 participants aged 18-39 years, 488 (90.7%) indicated that they are minimizing contact with family, friends, and coworkers, and 50 (9.3%) indicated that they are not. Of the 366 participants aged 40-64 years, 340 (92.9%) reported that they are minimizing contact with family, friends, and coworkers, and 26 (7.1%) reported that they are not. Of the 65 participants aged ≥65 years, 63 (96.9%) indicated that they are minimizing contact with family, friends, and coworkers, and 2 (3.1%) indicated that they are not.

Social distancing from family, friends, and coworkers was observed to be significantly associated with sex (*P*=.001). Of the 559 female participants, 528 (94.4%) indicated that they are minimizing contact with family, friends, and coworkers, and 31 (5.6%) indicated that they are not. Of the 410 male participants, 363 (88.5%) reported that they are minimizing contact with family, friends, and coworkers, and 47 (11.5%) reported that they are not.

CDC Guidelines Adherence

It was found that 21 (2.2%) of the 969 study participants were currently diagnosed with COVID-19 by a medical health professional. It was also found that 13 (61.9%) of these patients were prescribed and taking Plaquenil as a treatment. This is not a CDC guideline, and this was not associated with any such guidelines within the survey instrument. It has only been included here for reference, as several studies have shown potential efficacy of treatment with hydroxychloroquine, and
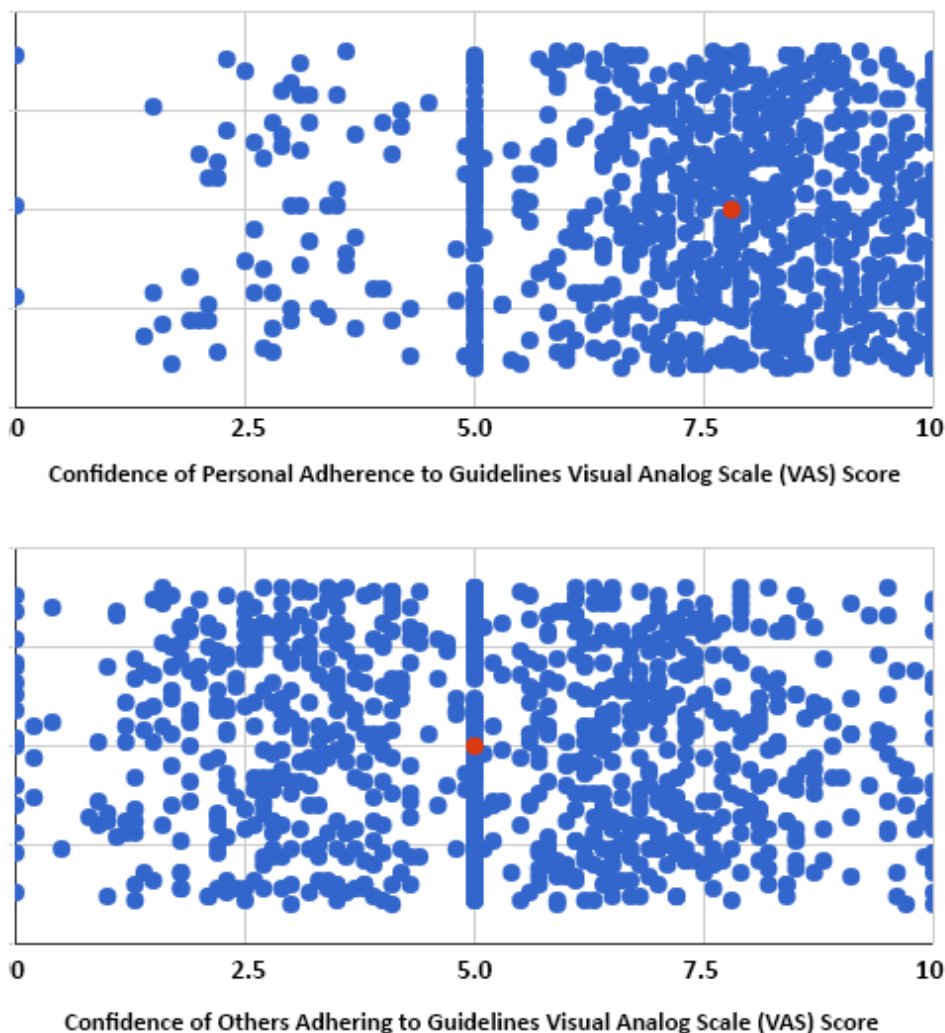
this information has been widely distributed by mainstream media [18,19].

Participants were asked to provide answers to a series of questions to determine if sanitary and social distancing guidelines were being followed as recommended by the CDC (as of March 24, 2020) [1]. It was found that, of the 969 participants, 92.3% (n=894) of individuals would stay home if they were sick unless they needed to seek medical care; 91.2% (n=884) would wear a face mask when around other people and before entering a health care provider's office; 85.8% (n=831) avoid touching their eyes, nose, and mouth with unwashed hands; 95.4% (n=924) indicated that they are avoiding close contact with people who are sick; 95.5% (n=925) throw used tissues in the trash; 92.6% (n=897) indicated they would wash hands with soap and water often for at least 20 seconds; and 87.8% (n=851) would use hand sanitizer that contains at least 60% alcohol if soap and water were not readily available. Over three-fourths (n=769, 79.4%) of study participants indicated that they wash their hands immediately after coughing or sneezing, and 73.7% (n=714) indicated that they clean and disinfect frequently touched surfaces daily. Additionally, 82.1% (n=796) clean surfaces that are visibly dirty with detergent or soap prior to disinfection.

A majority (n=615, 63.5%) of the 969 study participants indicated that they would not wear gloves or a face mask when going to public places if they were not sick. However, 50.2% (n=486) of study participants believe that the use of gloves or face masks by the general population helps prevent the spread of COVID-19.

Participants were asked to indicate their confidence, where 0 represented no confidence and 10 represented full confidence, regarding their personal ability to minimize becoming infected with COVID-19. The participants demonstrated a mean of 7.53 (95% CI 7.41-7.65). Regarding their confidence in other people following guidelines to prevent the likelihood of COVID-19 spread, participants reported a mean of 5.26 (95% CI 5.11-5.40), which was significantly lower when compared to participant confidence in their personal ability ($t_{1864.4}$=–23.572, *P*<.001). The majority of the 969 participants (n=717, 74%) believed that those younger than 40 years are contributing to the spread of COVID-19 more so than those 40 years or older (Figure 6).

Figure 6. Scatter plot depicting visual analog scores of all study participants, demonstrating the perception of confidence in personal ability to minimize the risk of infection, with a mean of 7.51 (SD 1.93) and a median of 7.80 (red dot), and confidence that others are adhering to the Centers for Disease Control and Prevention guidelines to minimize the likelihood of COVID-19 infection spread, with 0 representing no confidence and 10 the most confidence, with a mean of 5.28 (SD 2.35) and a median of 5.0 (red dot).

Confidence of Personal Adherence to Guidelines Visual Analog Scale (VAS) Score

Confidence of Others Adhering to Guidelines Visual Analog Scale (VAS) Score

Ethical Considerations

Study participants were asked several yes or no questions about ethical considerations with respect to resource management and the availability of medical care. When asked if they were buying or would buy excess amounts of sanitation supplies (ie, hand sanitizer, sanitizing wipes, and hand soaps) despite knowing there may not be enough supplies for everyone, 89 (9.2%) of the 969 indicated that they were or would, and 90.8% (n=880) reported that they were not or would not. When buying essential supplies, study participants (n=715, 73.8%) did not believe that people take the needs of others into consideration (Multimedia Appendix 5).

When asked if they would pay for medications they might need, even if they were not currently sick, 199 (20.5%) of the 969 participants indicated they would do so. Similarly, 171 (17.6%) reported that even if they were not sick, they would buy and use face masks with active knowledge of the limited supply available to health care workers and sick people. Although the

majority (n=775, 80.0%) indicated they were not upset when asked if they were upset that they could not see their physicians in person for reasons other than COVID-19–related concerns, knowing that in-person exposure would increase their risk of acquiring the virus, 194 (20.0%) reported they were upset they could not see their physician.

Travel Considerations

Study participants were asked additional yes or no questions about their thoughts regarding travel. The majority of the 969 responders indicated that they both have not traveled outside of the United States within the previous 3 months (n=901, 93.0%) and did not have plans to travel outside of the United States within the coming 3 months (n=917, 94.6%). Of the 52 (5.4%) participants who reported plans to travel internationally, 28 (53.8%) planned on cancelling their arrangements, though 24 (46.2%) did not. The majority of participants (n=558, 58.6%) believed that US citizens should be allowed to return to the United States, but 401 (41.4%) did not. The majority (n=651,

67.2%) indicated that non-US citizens currently in the United States should be allowed to return to their countries of origin.

The majority of the 969 participants (n=734, 75.9%) indicated that they did not plan on travelling domestically within the United States in the next 3 months. Of the 234 (24.1%) participants who reported planning on travelling domestically, 136 (58.1%) did not plan on cancelling their arrangements. The majority of participants (n=550, 56.8%) indicated that individuals and families should not be allowed to travel domestically within the United States. A total of 794 (81.9%) believed that those younger than 40 years should cancel travel plans, and 859 (88.6%) indicated that those 40 years or older should cancel travel plans. When asked how important travel is as it relates to the current COVID-19 pandemic, where 0 was of no importance at all and 10 was of the most importance, the mean response was 6.87 (95% CI 6.68-7.06).

Proactive Measures

Study participants were asked yes or no questions about several proactive measures they may or may not have taken regarding avoiding infection and stocking supplies. When asked if they have taken proactive measures to protect their health, the majority of the 969 participants responded "yes" (n=921, 95.0%), and 17 (1.75%) indicated that they have started taking Plaquenil as a preventative measure.

Of the 701 (72.3%) employed participants, 440 (62.8%) indicated that they were working from home. Of the 261 employed participants not working from home (37.2% of employed participants), the majority (n=199, 76.2%) stated they were not working from home because their employers had not made that option available (Multimedia Appendix 6).

The majority of the 969 respondents (n=737, 76.1%) indicated that they had stocked up on essential supplies such as food, sanitizing products, and necessary medications for chronic health conditions. The majority of respondents (n=693, 71.5%) believed that they had enough essential supplies for all individuals in their household, and 869 (89.7%) indicated that other people were proactively buying too many essential supplies. A total of 473 (48.8%) respondents believed that there are enough essential supplies for all individuals and families in the United States while 496 (51.2%) did not. The majority of respondents (n=715, 73.8%) did not believe individuals take the needs of others into account when purchasing essential supplies.

When asked in how many of the last 10 days had they been concerned that they would be unable to obtain resources such as food, sanitizing products, and necessary medications for themselves or their families, 230 (23.7%) of 969 participants responded that they had not been concerned, and 153 (15.8%) responded that they had been concerned every day in the past 10 days.

Discussion

Principal Results

The study showed several key differences in perception. Older individuals are at a higher risk of morbidity and mortality from COVID-19 infection, and their perception of the crisis was demonstrably more severe than that of the younger population (aged <40 years). Males were less concerned than females regarding the severity of the pandemic and were less likely to practice social distancing. Although there is no established difference in sex with regard to susceptibility to COVID-19 infection, there is limited epidemiological data to suggest increased burden of morbidity and mortality in males [20]. The role of differences in perceived severity of the pandemic and social distancing practices between males and females may contribute to this increased susceptibility because males may underestimate the threat of COVID-19 and engage less rigorously in social distancing.

The mental health implications of social distancing were greatest for younger people despite being at lower risk of morbidity and mortality as compared to older people. As social distancing has resulted in many social interactions becoming confined to the internet, it would be expected that younger people would be least affected by this transition, as younger generations are commonly more adept at using technology and traditionally more associated with the use of social media as a component of their social interactions. However, it may be possible that, because younger people experience a greater proportion of online interactions, they may place a larger relative value on in-person interactions than do older individuals. Because social distancing has resulted in a quick and dramatic reduction of in-person interactions, this may have disproportionately affected younger people. It is also possible that younger people are less likely to live with others when compared to those 40 years or older. This would mean that, during this period of social distancing when people are largely confined to their homes, young people are disproportionately isolated compared to older people. However, information regarding cohabitants was not collected in this survey. The larger negative mental health impact due to social distancing observed in younger people may also be due to their arguably greater potential loss of income. Interestingly enough, there was no significant difference observed between income groups, so financial factors alone cannot explain the fear and emotional toll that younger individuals are reportedly experiencing.

Though it was expected that high-risk individuals would perceive this pandemic as more severe than low-risk individuals, no significant difference was observed. Higher severity VAS scores were anticipated among high-risk participants because, by definition, they are at elevated risk for morbidity and mortality from infection. It is possible that a portion of high-risk individuals are not aware of their augmented risk or that a portion of low-risk individuals mistakenly believe they are at elevated risk. The collected data indicates that the public may be misevaluating the threat of morbidity and mortality of COVID-19. It is conceivable that this misalignment between perceived and actual threat may lead to noncompliance with guidelines, especially if one believes their risk to be less than it actually is. This may also lead to a disproportionate hoarding of resources among those at lower risk relative to those at higher risk and may have needless deleterious effect on mental health, especially if one believes their risk to be greater than it actually is.

The data collected here suggests that younger people believe the US government has responded to this crisis inappropriately while older people appear to be more supportive of the government's actions. However, the mean VAS response for all age groups was less than 6.0 (scale from 0-10), suggesting that approval of government response is low regardless of age. Although all age groups indicated concern regarding their personal finances, this concern did not differ between age groups. However, older individuals appear to be more concerned about the United States and global economy than are younger people.

Though many may think they know what practices to adopt to prevent viral spread, many participants said that if they were infected they would not wear a mask in public. This indicates that the public may be largely unaware of their lack in knowledge regarding the transmissibility of COVID-19 or that they are indifferent to contributing to the spread of COVID-19. Additionally, although many participants assumed that they knew how to minimize viral spread, they also assumed that others did not, and the degree to which they were confident in their own ability to reduce viral spread was significantly greater than the degree to which they believed other people could contribute to minimizing viral transmission. This may make disseminating accurate information and changing deleterious practices more difficult, as the data here suggests people are likely to think they are engaging in safe behavior when they might not be and that they believe their behavior is appropriate, while the actions of others are not.

Although hoarding practices are widely known to be disruptive to care, many admitted they would purchase unneeded materials and medicines despite the deleterious impact on public health. Furthermore, though the majority stated they have enough essential supplies for themselves and family, less than half believed there were enough essential resources for everyone, and the majority believed that people do not account for the needs of others when securing essential resources. This data suggests that possibly underlying hoarding behavior is a fear of the finiteness of resources and a lack of trust in others to responsibly ration.

Interestingly, though the spread of COVID-19 across boarders has been in part attributed to travel, of those with travel plans both foreign and domestic, 40%-60% planned to continue travelling. This may be indicative of a fatigue created by social distancing that, due to a newly imposed relative isolation, people are wanting to travel despite the risks of contracting and spreading viral infection. This may also reflect the need to travel to self-isolate with family and loved ones, rather than to do so alone. There may also be a nonnegligible portion of the population that remains unaware of the effect of travelling on viral spread.

Limitations

Although crowdsourcing is an extremely powerful tool, limitations exist. The potential limitation of using Amazon MTurk may be that study participants could submit multiple survey responses. Individuals could also circumvent the survey process completely by using a random number generator to create survey completion codes that are required for study participants to claim their wages. However, this did not have any impact on the overall study results, as all surveys were screened for completeness, quality, uniqueness, and accuracy, and duplicated responses were excluded. In addition, all completion codes were screened for authenticity, and time from survey start to survey completion was measured.

Inherent to many surveying methodologies, there also exists an internal bias with respect to those individuals who choose to take a survey about COVID-19. Those who have been impacted more severely or closely by the global implications of this pandemic may be more inclined to disseminate their opinions. The financial interest of the study participants may pose another potential limitation, yet the authors believe that this does not significantly skew the study results, with participants being compensated a total of US $0.25 for their response. In addition, the questions within the survey are not easily influenced by personal financial gain. However, MTurk remains a powerful platform in its ability to crowdsource an accurate representation of the US population as demonstrated by our demographic data being extremely similar to the data published in 2019 by the US Census Bureau. This has also been reflected by many prior crowdsourcing studies using the MTurk tool [21-25].

Conclusions

The authors believe that COVID-19 is contributing to tremendous emotional, financial, and health-related stress. This study helps to elucidate some of the public's fears about this disease as well as several key areas where better education may help improve outcomes. Namely, there needs to be a targeted effort to inform those that are infected about the importance of isolation and personal protective equipment. Furthermore, efforts should target hoarding practices and travel. Age-related differences in the perception of severity may change with time, and there is no good intervention to prevent these at this time. The emotional toll will be great, and the negative impact on mental health may continue to increase, especially in the younger population, though the long-term effects of this pandemic will be difficult to predict. Campaigns to target those in need of mental health resources now may curb the long-term negative mental health effects of this pandemic. In the future, this study will be repeated to evaluate how perception has changed over time and to better determine the long-term effects on the general population. Several efforts have already been made to continue to understand public opinions across race and ethnicity [26].

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Survey instrument used for data collection.
[PDF File (Adobe PDF File), 68 KB - jmir_v22i11e19768_app1.pdf]

Multimedia Appendix 2
CHERRIES (Checklist for Reporting Results of Internet E-Surveys).
[DOCX File , 18 KB - jmir_v22i11e19768_app2.docx]

Multimedia Appendix 3
Comparison of the distribution of ethnicity across all study participants (n=969) and the 2019 US census data published by the US Census Bureau.
[PNG File , 53 KB - jmir_v22i11e19768_app3.png]

Multimedia Appendix 4
Scatter plot depicting the annual household income of each individual study participant included in our analysis (n=969).
[PNG File , 101 KB - jmir_v22i11e19768_app4.png]

Multimedia Appendix 5
Answers to yes or no questions asked to all study participants (n=969) with respect to ethical considerations of stockpiling or hoarding materials and other societal considerations.
[PNG File , 331 KB - jmir_v22i11e19768_app5.png]

Multimedia Appendix 6
Answers to yes or no questions asked to all study participants (n=969) with respect to proactive measures taken in response to the COVID-19 pandemic.
[PNG File , 258 KB - jmir_v22i11e19768_app6.png]

References

1. Coronavirus (COVID-19). National Center for Health Statistics. 2019. URL: https://www.cdc.gov/nchs/covid19/index.htm
2. Velavan T, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280. [doi: 10.1111/tmi.13383] [Medline: 32052514]
3. Coronavirus disease 2019 (COVID-19) situation report – 59. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200319-sitrep-59-covid-19.pdf
4. Burke R, Midgley CM, Dratch A, Fenstersheib M, Haupt T, Holshue M, et al. Active monitoring of persons exposed to patients with confirmed COVID-19 - United States, January-February 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 06;69(9):245-246. [doi: 10.15585/mmwr.mm6909e1] [Medline: 32134909]
5. Garg VK, Bidani R, Rich EP, Hershey E, Hershey CO. Asthma patients' knowledge, perception, and adherence to the asthma guidelines. J Asthma 2005 Oct;42(8):633-638. [doi: 10.1080/02770900500263806] [Medline: 16266952]
6. Komajda M, Lapuerta P, Hermans N, Gonzalez-Juanatey JR, van Veldhuisen DJ, Erdmann E, et al. Adherence to guidelines is a predictor of outcome in chronic heart failure: the MAHLER survey. Eur Heart J 2005 Aug;26(16):1653-1659. [doi: 10.1093/eurheartj/ehi251] [Medline: 15827061]
7. Ung C, Zhang E, Alfaro T, Murakami Y, Zhang M, Seider MI, et al. Glaucoma severity and medication adherence in a county hospital population. Ophthalmology 2013 Jun;120(6):1150-1157 [FREE Full text] [doi: 10.1016/j.ophtha.2012.11.026] [Medline: 23453512]
8. Zapka JG, Stoddard A, Maul L, Costanza ME. Interval adherence to mammography screening guidelines. Med Care 1991 Aug;29(8):697-707. [doi: 10.1097/00005650-199108000-00003] [Medline: 1875738]
9. Subramanian S, Klosterman M, Amonkar MM, Hunt TL. Adherence with colorectal cancer screening guidelines: a review. Prev Med 2004 May;38(5):536-550. [doi: 10.1016/j.ypmed.2003.12.011] [Medline: 15066356]
10. Cunningham JA, Godinho A, Kushnir V. Using Mechanical Turk to recruit participants for internet intervention research: experience from recruitment for four trials targeting hazardous alcohol consumption. BMC Med Res Methodol 2017 Dec 01;17(1):156 [FREE Full text] [doi: 10.1186/s12874-017-0440-3] [Medline: 29191161]
11. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
12. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judgment Decision Making 2010;5(5):411-419.

13.   Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]

14.   Kelley PG. 2010 Presented at: SOUPS '10: Symposium on Usable Privacy and Security; 2010; Redmond, WA.

15.   Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

16.   Ipeirotis PG, Provost F, Wang J. Quality management on Amazon Mechanical Turk. 2010 Presented at: KDD '10: The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2010; Washington, DC. [doi: 10.1145/1837885.1837906]

17.   QuickFacts: United States. U.S. Census Bureau. 2019. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219

18.   Colson P, Rolain J, Lagier J, Brouqui P, Raoult D. Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. Int J Antimicrob Agents 2020 Apr;55(4):105932 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105932] [Medline: 32145363]

19.   Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H, et al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov 2020;6:16. [doi: 10.1038/s41421-020-0156-0] [Medline: 32194981]

20.   Cai H. Sex difference and smoking predisposition in patients with COVID-19. Lancet Respir Med 2020 Apr;8(4):e20 [FREE Full text] [doi: 10.1016/S2213-2600(20)30117-X] [Medline: 32171067]

21.   Shauly O, Gould DJ, Patel KM. Assessment of mental illness in patients seeking rhinoplasty. Plast Reconstr Surg 2019;7(8S-1):3. [doi: 10.1097/01.GOX.0000584200.93095.4f]

22.   Shauly O, Gould DJ, Patel KM. Crowdsourcing the public's perception of microsurgical and nonsurgical interventions for erectile dysfunction. Eur J Plast Surg 2020 Mar 11;43(4):443-452. [doi: 10.1007/s00238-020-01639-7]

23.   Gould D, Shauly O, Qureshi AA, Stevens WG. Defining "Ideal Abs" through a crowdsourcing-based assessment. Aesthet Surg J 2020 Mar 23;40(4):NP167-NP173. [doi: 10.1093/asj/sjz344] [Medline: 32022865]

24.   Shauly O, Gould DJ, Patel KM. The public's perception of interventions for migraine headache disorders: a crowdsourcing population-based study. Aesthetic Surg J Open Forum 2019:A. [doi: 10.1093/asjof/ojz007]

25.   Vartanian E, Gould DJ, Hammoudeh ZS, Azadgoli B, Stevens WG, Macias LH. The ideal thigh: a crowdsourcing-based assessment of ideal thigh aesthetic and implications for gluteal fat grafting. Aesthet Surg J 2018 Jul 13;38(8):861-869. [doi: 10.1093/asj/sjx191] [Medline: 29365056]

26.   Jones J, Sullivan PS, Sanchez TH, Guest JL, Hall EW, Luisi N, et al. Similarities and differences in COVID-19 awareness, concern, and symptoms by race and ethnicity in the United States: cross-sectional survey. J Med Internet Res 2020 Jul 10;22(7):e20001 [FREE Full text] [doi: 10.2196/20001] [Medline: 32614778]

Abbreviations

CDC: Centers for Disease Control and Prevention
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COPD: chronic obstructive pulmonary disease
MTurk: Mechanical Turk
REDCap: Research Electronic Data Capture
VAS: visual analog scale

XSL•FO
RenderX

information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO

RenderX

<u>Original Paper</u>

Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data

Junze Wang[1,2], PhD; Ying Zhou[1,2], BA; Wei Zhang[3], PhD; Richard Evans[4], PhD; Chengyan Zhu[5], PhD

[1]College of Public Administration, Huazhong University of Science and Technology, Wuhan, China

[2]Non-traditional Security Center, Huazhong University of Science and Technology, Wuhan, China

[3]School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

[4]College of Engineering, Design and Physical Sciences, Brunel University London, London, United Kingdom

[5]School of Political Science and Public Administration, Wuhan University, Wuhan, China

Corresponding Author:
Wei Zhang, PhD
School of Medicine and Health Management
Huazhong University of Science and Technology
13# Hangkong Road
Qiaokou District
Wuhan, 430074
China
Phone: 86 13397110378
Email: weizhanghust@hust.edu.cn

Abstract

Background: The COVID-19 pandemic has created a global health crisis that is affecting economies and societies worldwide. During times of uncertainty and unexpected change, people have turned to social media platforms as communication tools and primary information sources. Platforms such as Twitter and Sina Weibo have allowed communities to share discussion and emotional support; they also play important roles for individuals, governments, and organizations in exchanging information and expressing opinions. However, research that studies the main concerns expressed by social media users during the pandemic is limited.

Objective: The aim of this study was to examine the main concerns raised and discussed by citizens on Sina Weibo, the largest social media platform in China, during the COVID-19 pandemic.

Methods: We used a web crawler tool and a set of predefined search terms (*New Coronavirus Pneumonia*, *New Coronavirus*, and *COVID-19*) to investigate concerns raised by Sina Weibo users. Textual information and metadata (number of likes, comments, retweets, publishing time, and publishing location) of microblog posts published between December 1, 2019, and July 32, 2020, were collected. After segmenting the words of the collected text, we used a topic modeling technique, latent Dirichlet allocation (LDA), to identify the most common topics posted by users. We analyzed the emotional tendencies of the topics, calculated the proportional distribution of the topics, performed user behavior analysis on the topics using data collected from the number of likes, comments, and retweets, and studied the changes in user concerns and differences in participation between citizens living in different regions of mainland China.

Results: Based on the 203,191 eligible microblog posts collected, we identified 17 topics and grouped them into 8 themes. These topics were pandemic statistics, domestic epidemic, epidemics in other countries worldwide, COVID-19 treatments, medical resources, economic shock, quarantine and investigation, patients' outcry for help, work and production resumption, psychological influence, joint prevention and control, material donation, epidemics in neighboring countries, vaccine development, fueling and saluting antiepidemic action, detection, and study resumption. The mean sentiment was positive for 11 topics and negative for 6 topics. The topic with the highest mean of retweets was domestic epidemic, while the topic with the highest mean of likes was quarantine and investigation.

Conclusions: Concerns expressed by social media users are highly correlated with the evolution of the global pandemic. During the COVID-19 pandemic, social media has provided a platform for Chinese government departments and organizations to better understand public concerns and demands. Similarly, social media has provided channels to disseminate information about epidemic

prevention and has influenced public attitudes and behaviors. Government departments, especially those related to health, can create appropriate policies in a timely manner through monitoring social media platforms to guide public opinion and behavior during epidemics.

Introduction

On June 29, 2020, the World Health Organization (WHO) marked the six-month anniversary of the COVID-19 outbreak [1]. The first case of the unknown pneumonia strain, now known as COVID-19 and caused by SARS-CoV-2, was reported in Wuhan City, Hubei Province, on December 31, 2019, and has subsequently made a profound impact worldwide. After clusters of pneumonia cases were reported in Wuhan in early January 2020, laboratory analyses were conducted that led to the epidemic being identified as a new coronavirus, officially named SARS-CoV-2 by the WHO. By July 3, 2020, 83,545 confirmed cases were reported in Mainland China [2]. After the outbreak of COVID-19 in China, the disease quickly spread globally. At the end of August 18, 2020, a total of 214 countries and regions had reported confirmed cases of COVID-19 worldwide, with the total number of cases exceeding 21 million and the total number of deaths worldwide exceeding 760,000 [3]. The global average mortality rate of COVID-19 is approximately 0.3%-1.5%; however, the mortality rates of COVID-19 in countries such as the United States, Brazil, and Mexico are much higher. With the worldwide spread of COVID-19, the epidemic has gradually attracted widespread attention and discussion on social media platforms. This trend was particularly evident on Sina Weibo, one of the largest social media platforms in China.

Extant studies have demonstrated that timely understanding of public attitudes and demands plays an important role in responding to public crises [4,5]. Most social media platforms possess media-oriented features that are crucial to mediating information dissemination. Citizens can receive the latest fact-checked information provided by governments on social media in a timely manner, while governments can use the publicly available information released by citizens to better understand public attitudes, concerns, and demands [6-8]. Sina Weibo, with over 500 million users by May 2020, provides a variety of communication mechanisms for citizen interaction, allowing the Chinese public to share information and exchange opinions [9-11]. By analyzing the concerns raised by citizens about COVID-19 in microblog posts shared on Sina Weibo, governments can better understand public attitudes and demands [12,13] and clarify existing challenges faced by government departments and organizations when dealing with the pandemic. This research provides important insights and implications for policy makers, especially those working in public health departments. The results offer deeper understanding of public perception and highlight shortcomings in practice to better meet public needs.

A large body of literature has examined the role of social media in analyzing public behavior, attitudes, and responses during times of public crisis. Researchers have mainly focused on the platform functionality, user behavior characteristics, and use of social media during crises. As most platforms are freely available in the public domain [14], they have become widely adopted methods for citizens to stay connected, discuss concerns and opinions, and escape the monotony of lockdown during the COVID-19 pandemic [15]. Researchers agree that social media has become an important medium for information dissemination during the pandemic [16] and is playing a unique role in information sharing [17] and health care discussion [18]. Some scholars have explored user behavior and connection networks on social media platforms and believe that personality characteristics affect users' behavior [19], while networks comprise both positive and negative connections [20,21]. The unique position and roles of social media in response to public crises have also attracted major attention from researchers. They concur that social media platforms play important roles in crisis management, especially in terms of providing citizens with timely information [22] and reducing citizens' anxiety and fear [23]. Based on existing research, in this paper, we collect data from the leading Chinese microblogging platform, Sina Weibo, to analyze the main concerns expressed by citizens during the COVID-19 pandemic.

Methodology

Data Collection

Sina Weibo is the leading microblogging platform in China; it enables users to send and receive short character-limited posts and retrieve textual content by searching for specified keywords over a defined date range. Using this functionality, we collected shared microblog posts related to COVID-19 during the time period of December 1, 2019, to July 31, 2020. The Octopus web crawler tool was used to search for predefined keywords, including *New Coronavirus Pneumonia*, *Coronavirus*, *New Coronavirus*, and *COVID-19*. In addition to the textual content collected, we sourced the metadata for each microblogged post, including number of likes, number of comments, number of retweets, publishing time, and publishing location. The location information in the metadata refers to the registered address of the Sina Weibo account associated with a microblog post. To obtain these data, we used the advanced search functionality of Sina Weibo.

Data Preprocessing

In the Chinese language, there is no obvious separation between words. Therefore, for the purpose of completing the latent Dirichlet allocation (LDA) processing tasks, it was necessary to add obvious separators between words. We performed

Chinese segmentation on the textual content collected during data preprocessing. A widely used Chinese word segmentation tool, ictclas, was employed to divide microblog posts into groups of words separated by spaces. For example, the microblog post "JD announced the donation of 1 million medical masks and 60,000 pieces of medical supplies to Wuhan City in batches" was segmented into "JD / announced / the / donation of / 1 million medical masks / and / 60,000 pieces of / medical supplies / to / Wuhan City / in batches". In addition, we removed stop words [24] from the texts, similar to removing stop words such as "an" and "the" in English text analysis. Chinese stop words mainly fall into the following two categories: widely used vocabulary, such as me, you, some, and every day, and auxiliary words for mood, adverbs, prepositions, conjunctions, and other words that have no meaning by themselves, such as in, yes, so, and then [25]. In addition, punctuation and characters such as emojis were removed.

Data Analysis

We applied topic modeling by specifying the number of topics required by the LDA to separate the set of microblog posts into defined clusters [26]. LDA can be used to identify the most common topics in microblog posts shared on Sina Weibo. Topic modeling is an unsupervised machine learning technique that can identify clusters in a collection of documents (microblog posts in our case). In this study, we used the LDA algorithm from the LDA4j package. LDA4j is implemented in the Java language of the LDA algorithm, and the project can be downloaded for free from GitHub [27].

LDA is a widely used topic modeling algorithm [28]. According to the LDA model, a document (ie, the text of a microblog post) is a collection of vocabularies and may include multiple topics. The goal of the LDA model is to speculate on the distribution of topics based on a given document [29]. With LDA modeling, we can map the given documents to a fixed set of topics and capture the representative words for each topic. Then, the natural clusters in the microblog post dataset can be established.

To determine the appropriate number of LDA topics, we used the coherence score to draw judgments; this method has been proposed to be useful for selecting a suitable number of LDA topics [30]. Through continuous adjustment of the number of topics, we found that when the number of topics was 17, the consistency score reached its optimal value [31]. Therefore, we set the LDA model to separate the set of microblog posts into 17 clusters; the text sets are provided in Multimedia Appendix 1.

Subsequently, we conducted manual analysis and selected representative and high-proportion keywords from the top 30 keywords of each topic. Next, consensus was reached on the 17 topics and related keywords. Lastly, we used these keywords to classify the microblog posts; we also obtained the number of microblog posts under each topic and the proportion of each topic in all related microblogs. Examples of microblog posts for each topic are provided in Multimedia Appendix 2.

An example microblog is as follows:

> *The sudden outbreak of new coronal pneumonia has had an impact on the country's economic operation. It has brought a negative GDP growth and an increasingly complicated international and domestic environment. However, under the strong hedging of the counter-cyclical adjustment policy, the resumption of production has advanced rapidly, while the main economic indicators show a rebound in March, and the decline rate narrowed significantly.*

This microblog post can be categorized into two different topics: work and production resumption and economic shock.

We also performed other analyses of the collected data, such as sentiment analysis. The sentiment scores varied between –1.0 and 1.0, with –1.0 being the most negative text and 1.0 being the most positive text. In addition, we calculated the interaction rate of users for each topic by analyzing the average number of retweets, likes, and comments per topic. Finally, using the publishing time and location, we analyzed changes in user concerns based on the time period and differences in user engagement in discussions related to COVID-19 based on the Chinese region.

Results

Using the web crawler tool and predefined search terms, we obtained a total of 203,191 microblog posts from the Sina Weibo platform that were shared between December 1, 2019, and July 31, 2020.

Microblog Analysis

Topics

According to the results obtained from the LDA and the keywords involved in each topic, we were able to group the topics into 8 themes: (1) patient admission; (2) treatment and research; (3) treatment resources; (4) fighting the epidemic together; (5) work to restore order; (6) prevention and control measures of COVID-19; (7) domestic and overseas pandemic situation; and (8) impact of COVID-19. The corresponding terms for each topic are shown in Table 1; these keywords were also used as the criteria for topic classification.

Table 1. Topics covered in microblog posts and the representative terms corresponding to each topic.

Theme	Topics	Terms corresponding to each topic
Patient admission	Patients' outcry for help	*help, attention, diffusion, receiving*
Treatment and research	COVID-19 treatments	*discharged from the hospital, cure, treatment, rehabilitation, Chinese Medicine*
	Vaccine development	*vaccines, prevention, drugs, clinical trials*
Treatment resources	Medical resources	*medical treatment, doctor, medical care, nurse, mobile cabin, ward*
Fighting the epidemic together	Material donation	*mask, materials, donations*
	Fueling and saluting antiepidemic action	*work with one heart, saluting, unity is strength, responsibility, persevere, fight*
Work to restore order	Work and production resumption	*resumption of work, resumption of production, employment, operation*
	Study resumption	*return to school, student, school, high school, college entrance examination*
Prevention and control measures of COVID-19	Quarantine and investigation	*quarantine, contact, fever, 14 days, investigation*
	Joint prevention and control	*headquarters, work leading group, joint defense, joint control*
	Detection	*nucleic acid, positive, negative*
Domestic and overseas pandemic situation	Domestic epidemic	*Wuhan, Hubei, prison, Zhejiang, Shandong*
	Epidemics in neighboring countries	*Japan, South Korea, Tokyo, Russia*
	Epidemics in other countries worldwide	*America, Trump, Britain, Italy, India, Brazil, France*
	Epidemic statistics	*confirmed cases, new cases, cumulative cases, suspected cases*
Impact of COVID-19	Economic shock	*economy, influence, market, shock*
	Psychological influence	*hope, worry, fear, terrible*

Theme 1: Patient Admission

The topic contained in this theme is patients' outcry for help, which relates to patients who have or may have contracted COVID-19 and their treatment situation. The cries for help of patients who are awaiting treatment have attracted widespread attention, while the subsequent reception and treatment of these patients has also attracted significant attention.

Theme 2: Treatment and Research

This theme contains two topics. The first topic is COVID-19 treatments. In this topic, the recovery of patients with COVID-19 received widespread attention, such as the number of people discharged from hospital and the number of people who recovered. At the same time, treatment methods have attracted widespread public attention, such as treatments using Chinese medicine. The second topic in this theme is vaccine development. In this topic, research progress in relation to vaccines and their clinical trials has attracted much comment. Meanwhile, there has been widespread debate about whether vaccines can achieve the goal of preventing the virus.

Theme 3: Treatment Resources

The topic in this theme is medical resources. This topic focuses on core aspects of medical resources; medical staff and hospital wards, the allocation and integration of medical resources, and the establishment of temporary hospitals such as mobile cabin hospitals are all of general concern.

Theme 4: Fighting the Epidemic Together

This theme includes two topics. The first topic is material donation. In the early stages of fighting the epidemic in mainland China, masks and other antiepidemic materials were extremely scarce. This issue aroused widespread concern, and the donation of various antiepidemic materials became an active topic. The second topic is fueling and saluting antiepidemic action. This topic includes two aspects; the first aspect is the confidence and determination to beat the pandemic. Keywords such as *fighting* and *defeat* represent the general attitude of Sina Weibo users towards the fight against COVID-19. The second aspect, cooperation in the fight against the pandemic, is often discussed; key phrases such as *work with one heart* and *unity is strength* were frequently mentioned.

Theme 5: Work to Restore Order

As the spread of COVID-19 in mainland China is gradually being controlled, the restoration of production and resumption of normal ways of living is becoming an active topic. This topic includes two subtopics. The first subtopic relates to work and production resumption. The progress and arrangements of the resumption of work and production have aroused widespread concern and discussion. This work, which is aimed at restoring normal production and operational order, has received widespread support from the Chinese public. The second topic is study resumption, which mainly involves two aspects: first, the time when students of all ages will return to school, and second, the time for the national college entrance examination.

Theme 6: Prevention and Control Measures of COVID-19

This theme includes three topics. The first topic is quarantine and investigation, which includes two aspects: the investigation of symptoms, such as cough and fever, and quarantine periods. In particular, the 14-day quarantine period has become standard.

The second topic is joint prevention and control. Taking measures of joint prevention and control and establishing a headquarters are important actions taken by the Chinese government in response to the epidemic. This action is also key for the Chinese government to contain the epidemic in the short term. Correspondingly, it has also been widely recognized and is of great concern to the public.

The last topic is detection. The main content related to this topic is nucleic acid detection, which is an important way to establish which patients have COVID-10 or have been infected with SARS-CoV-2. The topic of detection has aroused widespread discussion.

Theme 7: Domestic and Overseas Pandemic Situation

Four topics are included in this theme. The first topic is domestic epidemic. On the one hand, because Hubei and Wuhan are the main battlefields in the fight against the epidemic in China, the epidemic situation in these regions has received a significant amount of attention. On the other hand, the spread of the epidemic in China is also of great concern to the public. Epidemics in neighboring countries is the second topic. The peak of this topic appeared in mid-to-late February 2020, when the epidemic began to spread from China to neighboring countries, such as Japan, South Korea, and Russia. This has aroused heated discussion among Sina Weibo users.

The third topic is epidemics in other countries worldwide. At the time when this topic appeared, the epidemic had spread globally; therefore, the topic included many countries and regions. The United States, as the country hit hardest by the epidemic, has attracted widespread attention from Sina Weibo microbloggers. In particular, US President Donald Trump's statement that if COVID-19 deaths in the United States could be controlled to less than 100,000, "we all together have done a very good job" [32] has been a subject of heated discussion. At the same time, other countries with serious numbers of cases, such as Italy, India, Brazil, and France, have caused extensive debates. The fourth topic is epidemic statistics. The epidemic data for COVID-19, such as the numbers of confirmed cases, new cases, and suspected cases, has received much attention and discussion; public attention to and discussion of the epidemic data have continued throughout all stages of the evolution of the epidemic. The epidemic statistics can be divided into two parts: domestic epidemic data and international epidemic data. Domestic epidemic data refers to case data released by the National Health Commission. For example, by the end of July 31, 2020, a total of 78,989 cases had been cured in China, and a total of 84,337 confirmed cases had been

reported. International epidemic data is published by governments worldwide [33]. For example, by the end of July 31, 2020, the number of daily diagnosed cases globally exceeded 289,000, totaling 17.4 million [34]. The United States alone added more than 70,000 diagnosed cases in one day [35].

Theme 8: Impact of COVID-19

This theme has two topics. The first is economic shock. This topic mainly refers to the negative impact of the epidemic on economies and markets worldwide. The second topic is psychological influence. On the one hand, the outbreak of the epidemic has delivered a substantial psychological blow to the public, and emotions such as worry and fear have spread among citizens. On the other hand, the public also hopes and believes that the fight against the epidemic will be successful, and they remain in an optimistic mood.

Results of the Proportional Analysis of the Main Topics

To understand the specific degrees of concern for the various topics identified, we established the number and proportion of each topic by counting the number of microblog posts per topic. For example, the proportion of each topic is the number of microblog posts shared about the topic divided by the total number of microblogs (N=203,191). The results of our analysis are shown in Table 2.

As shown in Table 2, statistics relating to the pandemic can be viewed as a metric to ascertain citizen understanding of the current situation during the epidemic. The topic of pandemic statistics has received significant attention since the initial outbreak and is ranked first, with 26.8% of posts (54,513/203,191). The domestic epidemic situation has always been the focus of public attention, especially when the epidemic in mainland China had not been fully controlled. The number of microblog posts on this topic accounted for 18.1% of the posts (54,513/203,191), ranking it second. As COVID-19 spread worldwide, the number of microblog posts referring to the pandemic in other countries grew consistently; this topic is ranked third with a percentage of 13.9% (28,329/203,191).

The treatment of COVID-19 raised high expectations and received attention from the public for a long period, with a proportion of 11.1% (22,725/203,191). Medical resources are particularly important in the fight against the epidemic, and this topic accounts for 10.9% of posts (22,246/203,191). The great impact of COVID-19 on the economy is closely related to the interests of the public; therefore, the topic of economic shock also received a high degree of attention, ranked sixth with 10.1% of posts (20,699/203,191). As one of the most important measures for controlling and preventing the spread of COVID-19, the topic of detection and quarantine has received considerable attention at all stages of the pandemic since its outbreak, accounting for 9.12% of total posts (18,541/203,191). The topic of patients' outcry for help also attracted widespread attention, accounting for 8.61% of total posts (17,513/203,191).

Table 2. Numbers of microblog posts on Sina Weibo related to each topic (N=203,191), n (%).

Rank	Topic	Microblog posts
1	Pandemic statistics	54,513 (26.8)
2	Domestic epidemic	36,936 (18.1)
3	Epidemics in other countries worldwide	28,329 (13.9)
4	COVID-19 treatments	22,725 (11.1)
5	Medical resources	22,246 (10.9)
6	Economic shock	20,699 (10.1)
7	Quarantine and investigation	18,541 (9.12)
8	Patients' outcry for help	17,513 (8.61)
9	Work and production resumption	15,914 (7.83)
10	Psychological influence	14,485 (7.12)
11	Joint prevention and control	12,457 (6.13)
12	Material donation	12,434 (6.11)
13	Epidemics in neighboring countries	11,995 (5.90)
14	Vaccine development	9428 (4.64)
15	Fueling and saluting anti-epidemic action	9413 (4.63)
16	Detection	8939 (4.39)
17	Study resumption	6889 (3.39)

An important measure to restore economic order in China is work and production resumption, which is closely related to citizens' lives. The percentage of microblog posts on this topic is 7.83% (15,914/203,191). The epidemic also greatly affected the public psychologically; therefore, the topic of psychological impact also received much attention, with a post percentage of 7.12% (14,485/203,191). Joint prevention and control and material donations have received a certain amount of attention. These two topics ranked 11th and 12th, with proportions of 6.13% (12,457/203,191) and 6.11% (12,434/203,191), respectively. The situation in neighboring countries also aroused comment from Sina Weibo users, although the duration was short-lived; accordingly, this topic ranks 13th, with 5.90% of posts (11,995/203,191).

The development of vaccines has also been discussed by the public, with related microblog posts accounting for 4.64% of posts (9428/203,191). The fight against COVID-19 is a national battle in which citizens fuel discussions on how to overcome the pandemic. The percentage of posts relating to this topic was 4.63% (9413/203,191). As an important means of identifying patients who are infected with COVID-19, detection has also received a certain degree of attention, accounting for 4.39% (8939/203,191). After the pandemic in Mainland China was controlled, study resumption was also mentioned; however, compared to other topics, the degree of concern in this area is relatively low, accounting for only 3.39% of total posts (6889/203,191).

Results of Users' Interaction and Sentiment Analysis

Based on the collected microblog metadata, including the numbers of retweets, comments, and likes, we were able to analyze the levels of interaction between users for each topic.

At the same time, we performed a sentiment analysis to obtain the sentiment value for each topic. Table 3 shows the average numbers of retweets, comments, and likes for each topic, as well as their sentiment values. The calculation formula used to determine the sentiment value is $(a - b)/(a - b)$, where a is the number of tweets expressing positive emotions and b is the number of tweets expressing negative emotions [36].

It should be noted that the data reported on in this article included microblogs posted by opinion leaders. The numbers of retweets, comments, and likes for these microblog posts were high. For example, for the post "Zhong Nanshan said that Lianhuaqingwen capsules were proved to be effective in the treatment of COVID-19", the numbers of retweets, comments, and likes were 13,599, 19,022, and 344,780, respectively. Therefore, the average numbers of retweets, comments, and likes in this paper are high.

From Table 3, it can be seen that domestic epidemic, material donation, and medical resources ranked as the top three topics in terms of the number of retweets, demonstrating that Sina Weibo users have a strong desire to share this information. By analyzing the number of likes, the mean was found to be relatively high for the following topics: quarantine and investigation, fueling and saluting antiepidemic action, material donation, and COVID-19 treatments. According to the sentiment value for each topic, the numbers of topics with positive and negative sentiment values were 11 and 6, respectively. Furthermore, for most topics, the sentiment value was low, which indicates that the number of microblog posts with positive emotions is almost the same as the number of microblog posts with negative emotions. However, negative emotions were obvious, except for those related to pandemic statistics.

Table 3. Results of user interaction and sentiment analysis for each topic.

Rank	Topic	Retweets (mean)	Comments (mean)	Likes (mean)	Sentiment value
1	Work and production resumption	8	12	87	0.16
2	Epidemics in neighboring countries	29	22	199	−0.006
3	Coronavirus treatments	26	26	244	−0.05
4	Vaccine development	24	16	199	0.08
5	Pandemic statistics	15	12	178	−0.11
6	Economic shock	18	10	97	0.051
7	Material donation	35	17	245	0.058
8	Domestic epidemic	71	20	184	0.024
9	Medical resources	34	35	241	0.031
10	Quarantine and investigation	33	24	280	−0.049
11	Psychological influence	29	29	217	0.041
12	Joint prevention and control	15	13	236	0.146
13	Study resumption	7	10	129	0.083
14	Fueling and saluting antiepidemic action	19	13	248	0.073
15	Epidemics in other countries worldwide	11	10	118	−0.117
16	Patients' outcry for help	22	18	224	−0.017
17	Detection	14	16	129	0.041

Results of Spatiotemporal Analysis of Microblog Posts

By using the publishing times of the microblog posts, we were able to analyze the change in users' concerns during different time periods. Similarly, we used the publishing location to analyze the differences in user engagement in discussions relating to COVID-19 between different regions in China. Figure 1 illustrates how the proportions of topics changed over the time period. The figure includes the names and proportions of the top four topics with the highest levels of attention from December 2019 to July 2020. Considering the different months, the top four topics constantly changed. From December 2019 to July 2020, there were 9 different topics in the top four, including pandemic statistics, domestic epidemic, epidemics in other countries worldwide, COVID-19 treatments, medical resources, economic shock, work and production resumption, psychological influence, and material donation. In December 2019 and January 2020, four topics received a high level of attention: domestic epidemic, pandemic statistics, psychological influence, and material donation. In February and March 2020, the topics of domestic epidemic and pandemic statistics received more attention than before; the numbers of microblog posts on these topics ranked first and second, respectively. Meanwhile,

the degrees of attention paid to the topics of medical resources and COVID-19 treatments increased, and they entered the top four topics. In April and May 2020, resumption of work and production and economic shock became active topics of discussion. At the same time, the topics of pandemic statistics and epidemics in other countries worldwide retained a high degree of attention. In June and July 2020, the most popular topics changed little compared with previous months. Topics such as pandemic statistics, epidemics in other countries worldwide, and economic shock also retained a high degree of attention during this period.

Figure 2 presents a geographical ranking of the number of microblog posts. The location information was obtained from the registration addresses of the Weibo accounts that posted the microblogs. Of the 203,191 microblog posts collected, 127,009 (62.5%) contained the publishing location. By analyzing the publishing location, we identified the six regions with the largest numbers of microblog posts. It can be seen from Figure 2 that a much higher number of posts originated from Hubei Province than from any of the other provinces. Henan Province and Guangdong Province were ranked second and third, respectively, with other provinces ranking below these.

Figure 1. Changes in the proportions of posts about the main topics from January to May.
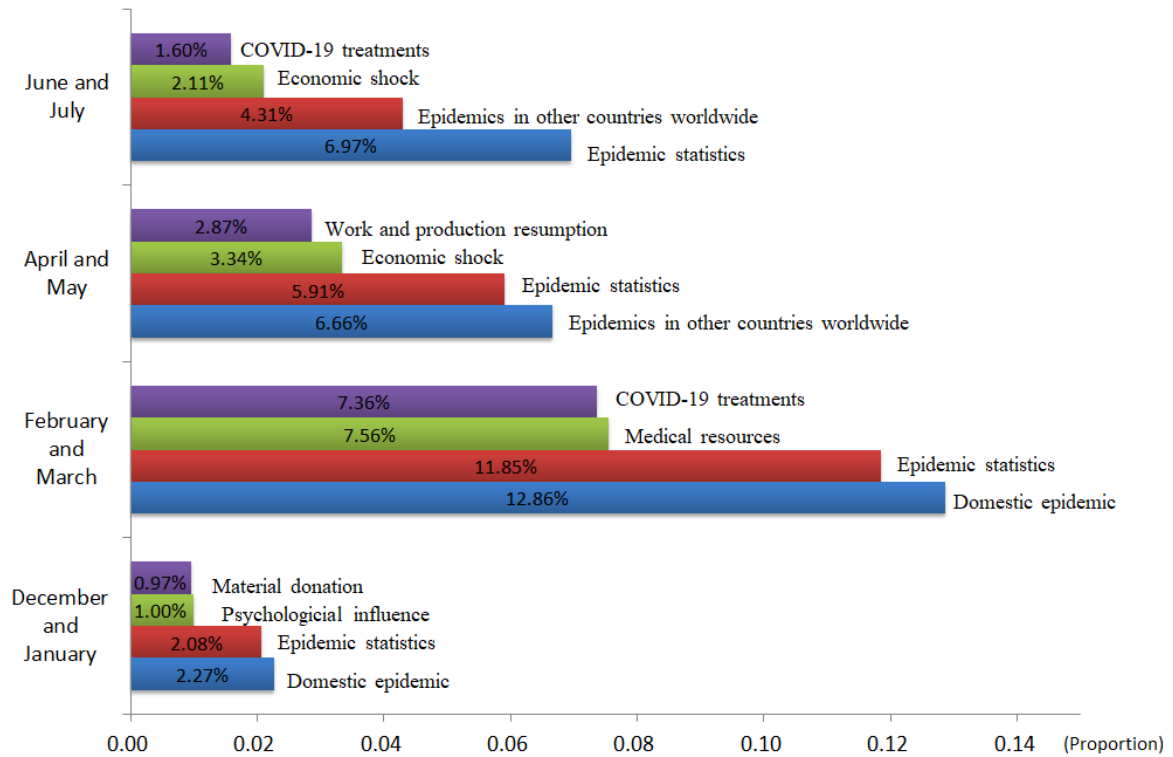
Figure 2. Numbers of microblog posts by Chinese province.

Discussion

Main Findings

Sina Weibo Users Mainly Focus on the Latest News About COVID-19 Treatment and the Status of the Pandemic Domestically and Globally

Our study, conducted from December 1, 2019, to July 31, 2020, identified 8 themes with 17 topics. Among these topics, content related to COVID-19 treatment aroused great attention, particularly content related to patients' outcry for help and emotions such as anxiety and panic. Due to the widespread negative emotions on Sina Weibo, misinformation and rumors can more readily influence public opinions, which negatively impacts antiepidemic actions. In addition, Sina Weibo users showed much interest in the progress of the treatment of COVID-19 and of vaccine development. At the same time, there has been considerable focus and discussion among a large number of Sina Weibo users on whether medical resources are sufficient and whether they can meet the treatment needs of patients.

Sina Weibo users also focused heavily on the real-time situation of the epidemic at home and abroad. In particular, many microblog posts contained real-time data on the development of the epidemic. Further, Sina Weibo users not only paid attention to the epidemic situation in the Chinese mainland but also greatly discussed the development and evolution of the epidemic in neighboring countries and the rest of the world. In addition, epidemic prevention and control measures, order restoration measures in the later stage of the epidemic, and the impact of the epidemic are included in the 17 topics that received attention from Sina Weibo users.

The Changes in the Number of Microblog Posts for the Various Topics in Different Time Periods Reflect the Process of the Occurrence, Development, and Demise of the COVID-19 Pandemic in China

Our findings show that December 2019 and January 2020 were the key periods of the response to the COVID-19 pandemic, with domestic epidemic and pandemic statistics receiving much attention during this period. At the same time, the sudden outbreak also affected the public psychologically. On the one hand, the public believes that the fight against the epidemic will eventually be won; on the other hand, emotions such as fear and worry are also spreading among the public.

February and March 2020 were critical periods for curbing the spread of the epidemic in mainland China [37]. During this period, pandemic statistics and the domestic epidemic were still causing high degrees of concern. At the same time, the surge in the number of patients caused the public to pay attention to COVID-19 treatments and medical resources. From April to May 2020, the pandemic in mainland China was basically under control; however, COVID-19 had begun to spread to other countries and regions. At this time, economic shock and work and production resumption became the most highly debated topics on Sina Weibo. Similarly, users paid great attention to the global spread of the pandemic and viewed real-time data.

From June to July 2020, the number of infections and deaths caused by COVID-19 worldwide continued to rise. The focus of Sina Weibo users shifted from China to the rest of the world, and the users began to pay greater attention to epidemics in other countries and the growing pandemic statistics.

The Location of Sina Weibo Users and Whether They Experienced Similar Events Are Closely Linked to Their Degree of Attention to Public Crises

According to the spatial distribution information collected on microblog posts, discussions related to COVID-19 differed among the various regions of China. The six regions with the highest degrees of participation were Hubei Province, Henan Province, Guangdong Province, Anhui Province, Jiangsu Province, and Hunan Province. Hubei Province is the epicenter of the COVID-19 epidemic in China, while Henan Province, Hunan Province, and Anhui Province are all adjacent to Hubei Province, with many citizens working in Hubei Province. Jiangsu Province, as an economically developed region, has a relatively high mobility rate of personnel. As a result, Sina Weibo users in these provinces were more likely to be concerned about the pandemic. Guangdong Province experienced the severe acute respiratory syndrome (SARS) epidemic in 2003, which is an important reason why more users in this region participated in discussions on the COVID-19 pandemic.

Concerns of Users of the Sina Weibo and Twitter Microblogging Platforms Show Similarities and Differences

A recent study reported the main topics related to COVID-19 that are of concern to Twitter users [36]. The authors analyzed 4 themes and 12 topics related to COVID-19 that Twitter users were concerned about from February 2 to March 15, 2020. They found that Twitter users were mainly concerned about the impact of COVID-19 on people and countries. For example, the number of deaths related to COVID-19 and the impact on citizens' emotions and psychology were mentioned in many tweets. In addition, the economic impact of COVID-19 was widely discussed. In particular, Twitter users mentioned two main methods for reducing the spread of COVID-19: masks and quarantine. Compared with these findings, Sina Weibo posts shared some similarities. Topics including economic impact and psychological influence received high attention on both Sina Weibo and Twitter [36]. However, at the same time, the main concerns of Sina Weibo users demonstrated unique characteristics. First, Twitter users focused on the causes and effects of the epidemic, while Sina Weibo users paid greater attention to prevention, control, and treatment. Various epidemic prevention and control measures, such as isolation, detection, and joint prevention and control, have attracted widespread attention among Sina Weibo users. Similarly, Sina Weibo users are highly concerned about content related to COVID-19 treatments, such as patients' outcry for help, medical resources, and treatment methods. In addition, citizens have continued to pay high degrees of attention to the development of the a vaccine against SARS-CoV-2. This may be due to the fact that the Chinese government has adopted a series of response measures and has made citizens aware of the severity and harmfulness of the pandemic; thus, Sina Weibo users are more concerned about

health-related topics such as the prevention, control, and treatment of the virus.

Second, Sina Weibo users focus on the real-time status of the epidemic. This includes not only the domestic epidemic situation, but also the development of the epidemic abroad. At the same time, in the middle and late stages of the epidemic, work to restore order, such as work and production resumption and study resumption, have attracted widespread attention. The following two reasons may have led to the emergence of the above unique concerns. First, as Chinese citizens were the first nation to experience the COVID-19 pandemic, they are more sensitive to the development and changes of the epidemic than people in other countries, and second, as the country that has demonstrated the most effective epidemic prevention and control, China has established conditions to restore order [38].

Prompt Guidance of Negative Emotions on Sina Weibo Is of Paramount Importance

The results of our sentiment analysis show that of the 17 topics, the numbers of topics with positive sentiment values and negative sentiment values are 11 and 6, respectively; for most topics, users have no obvious positive or negative emotional tendencies. However, it should be noted that there are still some topics in which the emotional value of a certain aspect is significant. For example, topics such as pandemic statistics demonstrated negative emotional tendencies. The continuous accumulation and spread of negative emotions on Sina Weibo may trigger irrational behavior among citizens, causing users to be affected by rumors or extreme emotions [19], such as group panic and denial of government support; thus, timely guidance to address negative emotions is essential.

Research Implications

With the outbreak and spread of COVID-19 worldwide, citizens have turned to social media channels, such as Sina Weibo, to share their opinions, seek clarity, and discuss topics related to the crisis. Previous studies have demonstrated that the analysis and control of public behavior and attitudes can effectively help governments cope in times of crisis [39]. As many countries start to experience a second wave of COVID-19 outbreaks, social media platforms can collect large amounts of information that reflect public behavior and attitudes. By analyzing this data, it is possible to identify the demands and behavioral characteristics of citizens [40].

In this paper, we collected data from Sina Weibo and analyzed it from the viewpoints of quantity, proportion, emotion, and space-time distribution. We identified the degrees of attention and the emotional tendencies of users toward various topics related to COVID-19, and we determined the time distribution for each topic and the spatial differences of users' participation in topic discussions. The results obtained enable deeper understanding of the views and attitudes of the public towards COVID-19, which is the premise of and basis for the prevention and control of this novel disease.

Strengths and Limitations

This paper is based on the existing theories of predecessors, combined with the timely topic of COVID-19. Through analysis of Sina Weibo microblogs, practical conclusions are drawn from the topics related to COVID-19 that are of concern to the public. At the same time, it is undeniable that this article also presents some limitations. First, the microblog posts collected in this study only include some posts related to COVID-19, not all of them. Therefore, the summary of the main concerns of Sina Weibo users is not comprehensive. Secondly, the majority of Sina Weibo users are young; therefore, the results of our analysis are more representative of younger citizens. Finally, although the number of Sina Weibo users is relatively large in China, we cannot draw a rushed conclusion that opinions expressed on the web represent public opinion in general.

Conclusions

The COVID-19 outbreak has had a crippling impact on the world economy and has presented numerous challenges to how people live and travel. Topics related to COVID-19 have attracted widespread attention and discussion on social media platforms, such as Sina Weibo. By analyzing the microblog posts of COVID-19 topics, we obtained the opinions of citizens on topics related to the virus which can lead to the identification of solutions to solve societal and economic problems [41]. In light of the comparative analysis of the main topics of Sina Weibo and Twitter, the topics of posts by Sina Weibo users show more regional characteristics. Judging from the distribution of microblog posts, the proportion of each topic reveals obvious differences. According to the sentiment analysis results, citizens reveal a positive emotional attitude. However, some negative emotions remain among the public, which requires timely guidance from Chinese governments. By analyzing the time distribution of microblog posts, it was observed that the proportion of topics changed significantly during different periods, while the main concerns of users changed with the development of the pandemic. Similarly, by analyzing the publishing location of microblog posts, we identified that user participation in the discussion of topics related to COVID-19 revealed obvious regional differences.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of microblogs for each topic.
[DOCX File , 16 KB - jmir_v22i11e22152_app1.docx]

Multimedia Appendix 2
Topics in Chinese.
[DOCX File , 47 KB - jmir_v22i11e22152_app2.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 29 June 2020. World Health Organization. 2020 Jun 29. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---29-june-2020 [accessed 2020-11-23]

2. The latest news about the COVID-19 by the time of 3rd July. Public Health Office, National Health Commission of China. 2020 Jul 03. URL: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd_13.shtml [accessed 2020-11-20]

3. Weekly Operational Update on COVID-19 21 August 2020. World Health Organization. 2020 Aug 21. URL: https://www.who.int/docs/default-source/coronaviruse/weekly-updates/wou-21082020.pdf?sfvrsn=e34ba8f3_2&download=true [accessed 2020-11-23]

4. Johnston KA, Taylor M, Ryan B. Emergency management communication: The paradox of the positive in public communication for preparedness. Public Relations Review 2020 Jun;46(2):101903. [doi: 10.1016/j.pubrev.2020.101903]

5. Huang Y, DiStaso M. Responding to a Health Crisis on Facebook: The Effects of Response Timing and Message Appeal. Public Relat Rev 2020 Sep;46(3):101909. [doi: 10.1016/j.pubrev.2020.101909]

6. Roy KC, Hasan S, Sadri AM, Cebrian M. Understanding the efficiency of social media based crisis communication during hurricane Sandy. Int J Inf Manage 2020 Jun;52:102060. [doi: 10.1016/j.ijinfomgt.2019.102060]

7. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]

8. Zhu C, Xu X, Zhang W, Chen J, Evans R. How Health Communication via Tik Tok Makes a Difference: A Content Analysis of Tik Tok Accounts Run by Chinese Provincial Health Committees. Int J Environ Res Public Health 2019 Dec 27;17(1):192 [FREE Full text] [doi: 10.3390/ijerph17010192] [Medline: 31892122]

9. Yan Q, Wu L, Zheng L. Social network based microblog user behavior analysis. Physica A 2013 Apr;392(7):1712-1723. [doi: 10.1016/j.physa.2012.12.008]

10. Zhang W, Deng Z, Evans R, Xiang F, Ye Q, Zeng R. Social Media Landscape of the Tertiary Referral Hospitals in China: Observational Descriptive Study. J Med Internet Res 2018 Aug 09;20(8):e249 [FREE Full text] [doi: 10.2196/jmir.9607] [Medline: 30093370]

11. Zhu C, Zeng R, Zhang W, Evans R, He R. Pregnancy-Related Information Seeking and Sharing in the Social Media Era Among Expectant Mothers: Qualitative Study. J Med Internet Res 2019 Dec 04;21(12):e13694 [FREE Full text] [doi: 10.2196/13694] [Medline: 31799939]

12. Bizid I, Nayef N, Boursier P, Doucet A. Detecting prominent microblog users over crisis events phases. Inf Sys 2018 Nov;78:173-188. [doi: 10.1016/j.is.2017.12.004]

13. Zhang W, Xu X, Zhang H, Chen Q. Online Participation Chaos: A Case Study of Chinese Government-Initiated E-Polity Square. Int J Public Sect Adm 2016 Jun 07;39(14):1195-1202. [doi: 10.1080/01900692.2015.1072218]

14. Lee PS, So CY, Lee F, Leung L, Chan M. Social media and political partisanship – A subaltern public sphere's role in democracy. Telemat Inform 2018 Oct;35(7):1949-1957. [doi: 10.1016/j.tele.2018.06.007]

15. Xie Y, Qiao R, Shao G, Chen H. Research on Chinese social media users' communication behaviors during public emergency events. Telemat Inform 2017 Jun;34(3):740-754. [doi: 10.1016/j.tele.2016.05.023]

16. Westerman D, Spence PR, Van Der Heide B. Social Media as Information Source: Recency of Updates and Credibility of Information. J Comput-Mediat Comm 2013 Nov 08;19(2):171-183. [doi: 10.1111/jcc4.12041]

17. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]

18. Osatuyi B. Information sharing on social media sites. Comput Hum Behav 2013 Nov;29(6):2622-2631. [doi: 10.1016/j.chb.2013.07.001]

19. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the Web?: The intersection of users' personality and social media use. Comput Hum Behav 2010 Mar;26(2):247-253. [doi: 10.1016/j.chb.2009.09.003]

20. Leskovec J, Huttenlocher D, Kleinberg J. Signed Networks in Social Media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Apr Presented at: CHI 2010: The SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA p. 1361-1370. [doi: 10.1145/1753326.1753532]

21. Zhang W, Deng Z, Hong Z, Evans R, Ma J, Zhang H. Unhappy Patients Are Not Alike: Content Analysis of the Negative Comments from China's Good Doctor Website. J Med Internet Res 2018 Jan 25;20(1):e35 [FREE Full text] [doi: 10.2196/jmir.8223] [Medline: 29371176]

XSL•FO
RenderX

22. Yin J, Lampert A, Cameron M, Robinson B, Power R. Using Social Media to Enhance Emergency Situation Awareness. IEEE Intell Syst 2012 Nov;27(6):52-59. [doi: 10.1109/mis.2012.6]

23. Simon T, Goldberg A, Adini B. Socializing in emergencies—A review of the use of social media in emergency situations. Int J Inf Manage 2015 Oct;35(5):609-619. [doi: 10.1016/j.ijinfomgt.2015.07.001]

24. Hajjem M, Latiri C. Combining IR and LDA Topic Modeling for Filtering Microblogs. Procedia Comput Sci 2017;112:761-770. [doi: 10.1016/j.procs.2017.08.166]

25. Rani R, Lobiyal D. Performance evaluation of text-mining models with Hindi stopwords lists. J King Saud Univ, Comp & Info Sci . [doi: 10.1016/j.jksuci.2020.03.003]

26. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]

27. LDA4j. GitHub. URL: https://github.com/hankcs/LDA4j [accessed 2020-11-20]

28. Arora S, Ge R, Halpern Y, Mimno D, Moitra A, Sontag D, et al. A Practical Algorithm for Topic Modeling with Provable Guarantees. ArXiv. Preprint posted online on Dec 19, 2012.

29. Pavlinek M, Podgorelec V. Text classification method based on self-training and LDA topic models. Expert Syst Appl 2017 Sep;80:83-93. [doi: 10.1016/j.eswa.2017.03.020]

30. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring Topic Coherence over Many Models and Many Topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.: Association for Computational Linguistics; 2012 Jul Presented at: 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; July 12-14, 2020; Jeju Island, Korea p. 952-961 URL: https://www.aclweb.org/anthology/D12-1087.pdf

31. Amoualian H, Lu W, Gaussier E, Balikas G, Amini M, Clausel M. Topical Coherence in LDA-based Models through Induced Segmentation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).: Association for Computational Linguistics; 2017 Jul Presented at: 55th Annual Meeting of the Association for Computational Linguistics; July 20-August 4, 2017; Vancouver, BC p. 1799-1809. [doi: 10.18653/v1/p17-1165]

32. Bohn K. Trump concedes US coronavirus death toll could be 100,000 or more. CNN. 2020 Mar 30. URL: https://www.cnn.com/2020/03/29/politics/trump-deaths-coronavirus/index.html [accessed 2020-11-23]

33. Chinese mainland reports 45 new COVID-19 cases. CGTN. 2020 Aug 01. URL: https://news.cgtn.com/news/2020-08-01/Chinese-mainland-reports-45-new-COVID-19-cases-SAOFIVVkL6/index.html [accessed 2020-11-23]

34. Coronavirus disease (COVID-19) Situation Report – 193. World Health Organization. 2020 Jul 31. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200731-covid-19-sitrep-193.pdf?sfvrsn=42a0221d_4 [accessed 2020-11-23]

35. Shumaker L. U.S. shatters coronavirus record with over 77,000 cases in a day. Reuters. 2020 Jul 16. URL: https://www.reuters.com/article/us-health-coronavirus-usa-records-idUSKCN24I014 [accessed 2020-11-23]

36. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

37. White paper - Fighting Covid-19: China in Action. China Daily. 2020 Jun. URL: https://covid-19.chinadaily.com.cn/a/202006/08/WS5edd8bd6a3108348172515ec.html [accessed 2020-11-20]

38. Shaoyong C, Huanming Z, Min L, Shuanghuan L. Comparision of microblogging service between Sina Weibo and Twitter. In: Proceedings of the 2011 International Conference on Computer Science and Network Technology.: IEEE; 2011 Presented at: 2011 International Conference on Computer Science and Network Technology; Harbin, China; December 24-26, 2011 p. 2259-2263. [doi: 10.1109/iccsnt.2011.6182424]

39. Xiaomei Z, Jing Y, Jianpei Z, Hongyu H. Microblog sentiment analysis with weak dependency connections. Knowl-Based Syst 2018 Feb;142:170-180. [doi: 10.1016/j.knosys.2017.11.035]

40. Fersini E, Pozzi FA, Messina E. Approval network: a novel approach for sentiment analysis in social networks. World Wide Web 2016 Oct 11;20(4):831-854. [doi: 10.1007/s11280-016-0419-8]

41. Saroj A, Pal S. Use of social media in crisis management: A survey. Int J Disaster Risk Reduct 2020 Sep;48:101584. [doi: 10.1016/j.ijdrr.2020.101584]

Abbreviations

LDA: latent Dirichlet allocation
SARS: severe acute respiratory syndrome

XSL·FO
RenderX

Original Paper

Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings

Jeffrey V Lazarus[1], DPhil; Adam Palayew[2]; Lauge Neimann Rasmussen[3]; Tue Helms Andersen[3]; Joey Nicholson[4]; Ole Norgaard[3]

[1]Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain

[2]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[3]Danish Diabetes Knowledge Center, Steno Diabetes Center Copenhagen, Gentofte, Denmark

[4]NYU Langone Health, NYU Grossman School of Medicine, NYU Health Sciences Library, New York, NY, United States

Corresponding Author:
Jeffrey V Lazarus, DPhil
Barcelona Institute for Global Health (ISGlobal)
Hospital Clínic
University of Barcelona
Carrer del Rosselló, 132
Barcelona
Spain
Phone: 34 608703573
Email: jeffrey.lazarus@isglobal.org

Abstract

Background: Since it was declared a pandemic on March 11, 2020, COVID-19 has dominated headlines around the world and researchers have generated thousands of scientific articles about the disease. The fast speed of publication has challenged researchers and other stakeholders to keep up with the volume of published articles. To search the literature effectively, researchers use databases such as PubMed.

Objective: The aim of this study is to evaluate the performance of different searches for COVID-19 records in PubMed and to assess the complexity of searches required.

Methods: We tested PubMed searches for COVID-19 to identify which search string performed best according to standard metrics (sensitivity, precision, and F-score). We evaluated the performance of 8 different searches in PubMed during the first 10 weeks of the COVID-19 pandemic to investigate how complex a search string is needed. We also tested omitting hyphens and space characters as well as applying quotation marks.

Results: The two most comprehensive search strings combining several free-text and indexed search terms performed best in terms of sensitivity (98.4%/98.7%) and F-score (96.5%/95.7%), but the single-term search COVID-19 performed best in terms of precision (95.3%) and well in terms of sensitivity (94.4%) and F-score (94.8%). The term Wuhan virus performed the worst: 7.7% for sensitivity, 78.1% for precision, and 14.0% for F-score. We found that deleting a hyphen or space character could omit a substantial number of records, especially when searching with SARS-CoV-2 as a single term.

Conclusions: Comprehensive search strings combining free-text and indexed search terms performed better than single-term searches in PubMed, but not by a large margin compared to the single term COVID-19. For everyday searches, certain single-term searches that are entered correctly are probably sufficient, whereas more comprehensive searches should be used for systematic reviews. Still, we suggest additional measures that the US National Library of Medicine could take to support all PubMed users in searching the COVID-19 literature.

XSL·FO
RenderX

Introduction

Since it was declared a pandemic on March 11, 2020, COVID-19 has dominated headlines around the world and generated thousands of scientific articles [1]. The fast speed of publication has challenged researchers and other stakeholders to keep up with the volume of published articles on the topic [2]. To search the literature, researchers and others use databases of peer-reviewed scientific articles. These databases require indexing and curation of articles, which is a time-consuming task; however, the investment in curated databases aims to establish reliable and more efficient article searching [3].

PubMed is the database of choice for many clinicians and researchers due to its ease of use, reputation, large number of indexed journals, and free access [4]. It is maintained by the US National Library of Medicine (NLM) [4-6]. PubMed provides subject filters to facilitate searching specific topics but as they are only reviewed once a year, it is unknown if a COVID-19 subject filter is under consideration [7]. However, at the beginning of the pandemic, NLM introduced a one-click search option on their dedicated website for COVID-19/SARS-CoV-2 [8]. By clicking on a button, PubMed users can quickly apply a ready-made search to identify potentially relevant COVID-19 records in PubMed. Searching structured databases like PubMed is generally considered to be most skillfully executed by employing both Medical Subject Headings (MeSH) and free-text terms [9-11]. However, conducting systematic, comprehensive, and transparent searches takes time and skill [12-14]. Thus, PubMed's one-click search is potentially a valuable shortcut for users worldwide dealing with the COVID-19 pandemic.

A recent survey showed that authors of systematic reviews have different perspectives on what constitutes an "effective" literature search [15]. Based on the survey, Cooper et al [15] note that review authors "appeared to locate effectiveness within a binary classification between types of review which are either comprehensive or non-comprehensive." In short, their article indicates that those valuing comprehensive literature searches tend to be literature searchers (eg, information specialists or librarians) who emphasize sensitivity as an evaluation metric for effectiveness and highlight Cochrane-style systematic reviews, with their ambition to identify all available evidence as an ideal. On the other hand, researchers and health professionals screening the literature tend to emphasize less comprehensive searching and value precision as an evaluation metric. In addition, they care about outcomes like the workload, time, and resources that are needed to handle identified records. However, most PubMed users only browse the first 20 records of their search results and refine their searches to make the number of hits manageable [3]. For such reasons, the NLM also has an algorithm that is designed to sort search results according to their relevance via the Best Match sorting option [16] and offers one-click filters (eg, publication date and article type).

These user behaviors and options reflect a different view on literature searches than those voiced by users conducting formal literature reviews [15]. This might be because PubMed users also consist of clinicians, health professionals, and other decision makers that search the literature not to conduct reviews but with other types of evidence use in mind.

No matter one's perspective, the effectiveness of any literature search depends on the semantic variability related to the question that the search is to inform. When a research topic is in flux, establishing common terminology is crucial to identify relevant records. A retrospective study of searching PubMed during the first 10 weeks of the 2009 H1N1 influenza pandemic showed that inconsistent disease naming and a lag in indexing increased the risk of missing relevant studies when searching the scientific literature during the 2009 pandemic [17]. The authors of the study recommended that, at the start of a pandemic, "the international scientific community should agree on nomenclature and the specific name to be used earlier, and the U.S. National Library of Medicine and other database providers should incorporate that in their indexing of all relevant articles" [17]. Regarding the COVID-19 pandemic, the naming of the disease occurred relatively early. On February 11, 2020, the International Committee on Taxonomy of Viruses officially named the novel coronavirus SARS-CoV-2 and the World Health Organization (WHO) named the disease COVID-19 [18]. However, article authors still use several other terms for the virus and the disease, such as simply "coronavirus" and, earlier, "Wuhan coronavirus" or "Wuhan virus."

In this study, we aimed to evaluate the performance of different searches for COVID-19 records in PubMed and to assess the complexity of searches that was required. Therefore, we compared the performance of PubMed's one-click search option with both simpler and more complex search strings for the first 10 weeks of the COVID-19 pandemic. We also examined the deletion of hyphens or spaces as well as quotation marks from the simple searches to mimic potential user errors or preferences, such as variant spellings of words and the tendency to use only simple keyword searches [19]. Finally, we discuss the searches in relation to the varying perspectives on search effectiveness outlined above to make recommendations on how NLM can improve PubMed.

Methods

General Methodology

We constructed a comprehensive COVID-19 search string and compared it to seven other related search strings (Table 1). We queried PubMed for each of these different searches and calculated their sensitivity, precision, and F-score using a COVID-19 database (LitCovid) maintained and curated by an NLM branch as a gold standard [20]. We then used these calculated values to compare the performance of the different search strings.

Table 1. Search strings and translations of the different searches.

Search title	Entered search	Translation in PubMed
Our comprehensive search (Search 1)	("COVID-19"[nm] OR "COVID-19 diagnostic testing"[nm] OR "COVID-19 drug treatment"[nm] OR "COVID-19 serotherapy"[nm] OR "COVID-19 vaccine"[nm] OR "severe acute respiratory syndrome coronavirus 2"[nm] OR 2019-ncov*[tiab] OR 2019ncov*[tiab] OR 2019-novel-cov*[tiab] OR coronavirus[ti] OR coronavirus-2*[tiab] OR coronavirus-disease-19*[tiab] OR corona-virus-disease-19*[tiab] OR coronavirus-disease-20*[tiab] OR corona-virus-disease-20*[tiab] OR covid-19*[tiab] OR covid19*[tiab] OR covid-20*[tiab] OR covid20*[tiab] OR ncov-2019*[tiab] OR ncov2019*[tiab] OR new-coronavirus[tiab] OR new-corona-virus[tiab] OR novel-coronavirus[tiab] OR novel-corona-virus[tiab] OR sars-2*[tiab] OR sars2*[tiab] OR sars-cov-19*[tiab] OR sars-cov19*[tiab] OR sarscov19*[tiab] OR sarscov-19*[tiab] OR sars-cov-2*[tiab] OR sars-cov2*[tiab] OR sarscov2*[tiab] OR sarscov-2*[tiab] OR (("Coronavirus"[mh] OR "Coronavirus Infections"[mh] OR betacoronavirus[tiab] OR beta-coronavirus[tiab] OR beta-corona-virus[tiab] OR corona-virus[tiab] OR coronavirus[tiab] OR sars*[tiab] OR severe-acute-respiratory*[tiab]) AND (2019[tiab] OR 2020[tiab] OR wuhan*[tiab] OR hubei*[tiab] OR china*[tiab] OR chinese*[tiab] OR outbreak*[tiab] OR epidemic*[tiab] OR pandemic*[tiab]))) AND 2019/12:3000[dp]	None
Shokraneh's [21] comprehensive search (Search 2)	((((((((((((((((((((("Betacoronavirus"[MeSH Terms] OR "Coronavirus Infections"[MeSH Terms]) OR "COVID-19"[Supplementary Concept]) OR "Coronavirus"[MeSH Terms]) OR "Severe Acute Respiratory Syndrome Coronavirus 2"[Supplementary Concept]) OR "2019nCoV"[All Fields]) OR "betacoronavirus*"[All Fields]) OR "corona virus*"[All Fields]) OR "coronavirus*"[All Fields]) OR "coronovirus*"[All Fields]) OR "CoV"[All Fields]) OR "CoV2"[All Fields]) OR "COVID"[All Fields]) OR (("COVID-19"[Supplementary Concept] OR "COVID-19"[All Fields]) OR "covid19"[All Fields])) OR (((((("COVID-19"[All Fields] OR "covid 2019"[All Fields]) OR "Severe Acute Respiratory Syndrome Coronavirus 2"[Supplementary Concept]) OR "Severe Acute Respiratory Syndrome Coronavirus 2"[All Fields]) OR "2019 ncov"[All Fields]) OR "SARS CoV 2"[All Fields]) OR "2019nCoV"[All Fields]) OR (("wuhan"[All Fields] AND ("Coronavirus"[MeSH Terms] OR "Coronavirus"[All Fields])) AND (2019/12/1:2019/12/31[Date - Publication] OR 2020/1/1:2020/12/31[Date - Publication]))))) OR "HCoV-19"[All Fields]) OR "nCoV"[All Fields]) OR "SARS CoV 2"[All Fields]) OR "SARS2"[All Fields]) OR "SARSCoV"[All Fields]) OR ((((("sars virus"[MeSH Terms] OR ("sars"[All Fields] AND "virus"[All Fields])) OR "sars virus"[All Fields]) OR ("sars"[All Fields] AND "CoV"[All Fields])) OR "sars cov"[All Fields])) OR (("Severe Acute Respiratory Syndrome Coronavirus 2"[Supplementary Concept] OR "Severe Acute Respiratory Syndrome Coronavirus 2"[All Fields]) OR "SARS CoV 2"[All Fields])) OR "severe acute respiratory syndrome cov*"[All Fields]) AND (2019/11/17:3000/12/31[Date - Entry] OR 2019/11/17:3000/12/31[Date - Publication])	None

Search title	Entered search	Translation in PubMed
One-click search (Search 3)	((wuhan[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND 2019/12[PDAT] : 2030[PDAT]) OR 2019-nCoV[All Fields] OR 2019nCoV[All Fields] OR COVID-19[All Fields] OR SARS-CoV-2[All Fields]	(((((("wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND 2019/12/1:2030/12/31[Date - Publication]) OR (("severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields]) OR "2019 ncov"[All Fields])) OR "2019nCoV"[All Fields]) OR ((((((("covid 19"[All Fields] OR "covid 2019"[All Fields]) OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept]) OR "severe acute respiratory syndrome coronavirus 2"[All Fields]) OR "2019 ncov"[All Fields]) OR "sars cov 2"[All Fields]) OR "2019nCoV"[All Fields]) OR (("wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND (2019/12/1:2019/12/31[Date - Publication] OR 2020/1/1:2020/12/31[Date - Publication])))) OR (("severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields]) OR "sars cov 2"[All Fields])
Single-term search for COVID-19 (Search 4)	COVID-19	"COVID-19"[All Fields] OR "COVID-2019"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV-2"[All Fields] OR "2019nCoV"[All Fields] OR (("Wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND (2019/12[PDAT] OR 2020[PDAT]))
Single-term search for SARS-CoV-2 (Search 5)	SARS-CoV-2	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "sars cov 2"[All Fields]
Single-term search for Coronavirus (Search 6)	Coronavirus	"coronavirus"[MeSH Terms] OR "coronavirus"[All Fields] OR "coronaviruses"[All Fields]
Single-term search for Wuhan coronavirus (Search 7)	Wuhan coronavirus	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "wuhan coronavirus"[All Fields]
Single-term search for Wuhan virus (Search 8)	Wuhan virus	("Wuhan"[All Fields] AND ((((((("virology"[MeSH Subheading] OR "virology"[All Fields]) OR "viruses"[All Fields]) OR "viruses"[MeSH Terms]) OR "virus s"[All Fields]) OR "viruse"[All Fields]) OR "virus"[All Fields]))

Constructing a Comprehensive Search String

A comprehensive search string (Search 1) was initially developed by ON, LNR, and THA to monitor newly published COVID-19–related studies as part of their work at the Danish Diabetes Knowledge Center. Subsequently, the search string was revised based on the authors' subject knowledge, analyses of free-text words and MeSH terms conducted in PubReMiner [22], and search strings developed by information specialists that were made publicly available (eg, by the Medical Library Association and different university libraries). Several versions were tested and reiterated before the final version was reached [14]. The final search string was then reviewed by AP and JVL.

Comparing Search Strings

In addition to Search 1, we conducted Searches 2 through 8, which included an additional comprehensive search developed by Shokraneh (Search 2) [21], the one-click search option developed by NLM (Search 3), as well as five common terms used to search PubMed for COVID-19–related records (Searches 4-8), to compare different comprehensive searches and compare the comprehensive searches to the simple PubMed queries [21] (Table 1). The automatic term-mapping feature in PubMed translates some of these basic queries to more comprehensive search strings that include synonyms and MeSH terms, as shown in Table 1. We observed that the COVID-19 translation does

not seem to follow the standard mapping process for automatic term mappings [23].

Simulating Historical Weekly Searches

We searched from March 11 to May 19, 2020, spanning a total of 10 weeks, or 70 days, from when the WHO declared COVID-19 a pandemic. We limited the searches to find only records registered in PubMed during each of the 10 weeks (Wednesday through Tuesday for each week). For each of the eight searches, we recorded and analyzed the weekly number of records by using the date of the records' entry to the PubMed database (EDAT field in PubMed). All searches were conducted in the current standard version of PubMed on June 26, 2020.

Analysis of the Searches

We compared the evaluation metrics sensitivity, precision, and F-score for each search. Sensitivity is a measure of search effectiveness and is defined as the number of relevant records in the gold standard data set identified by the search (true positives) over all relevant records in the gold standard data set (true positives and false negatives) [24]. Sensitivity is also known as recall; however, to describe literature search effectiveness, the term sensitivity is widely used (eg, by NLM when reporting the effectiveness of PubMed Clinical Queries Filters) [24,25]. Precision is a measure of search efficiency and is defined as the number of relevant records identified by the search (true positives) over the total number of identified records (true positives and false positives). Finally, the F-score is defined as the harmonic mean of the sensitivity and the precision (Table 2). We used the LitCovid database as the gold standard to compare the PubMed searches against [20]. This database contains COVID-19–related records solely from PubMed and is curated by the NLM based on daily broad PubMed searches. As we do not know which records were deemed not relevant by the NLM, we do not know the number of true negatives. Thus, it was not possible to reliably calculate other relevant metrics, such as the specificity and accuracy of the evaluated search strings. For the calculations of the evaluation metrics, the searches were rerun, limiting the dates from January 17, 2020 (the earliest date of any record in LitCovid), to May 19, 2020 (the end of the study period). For this period, the LitCovid database contains 14,018 records.

Table 2. Descriptions and calculations of metrics.

Metric	Description	Calculation[a]
Sensitivity	The probability that the search identified a record as relevant given that the record is relevant; also known as recall.	$T^+D^+/(T^+D^++T^-D^+)$
Precision	The probability that the record is relevant given that the record was identified as relevant; also known as positive predictive value (PPV). The number needed to read (NNR) can also be calculated by 1/precision.	$T^+D^+/(T^+D^++T^+D^-)$
F-score	The harmonic mean of the sensitivity and precision.	$2 \times$ sensitivity $\times$ precision/(sensitivity + precision)

[a]We denote T^+ as in the search string as compared to the LitCovid database, T^- as not in the search string as compared to the LitCovid database, D^+ as present in the LitCovid database as compared to the search string, and D^- as not present in the LitCovid database as compared to the search string.

Sensitivity to the Deletion or Addition of Hyphens, Spaces, and Quotation Marks

Spelling mistakes have previously been documented to affect PubMed results [26]. Further, users of PubMed may have different writing style preferences and choose to apply or leave out hyphens and spaces. To investigate the possible implications of entering different versions of the search terms COVID-19 (Search 4) and SARS-CoV-2 (Search 5), we compared the results after omitting a hyphen and/or a space (eg, COVID19, COVID 19, COVID-19). We also assessed the implications of surrounding search terms with quotation marks, assuming that some users may do this to run a highly specific search (eg, "COVID-19," "COVID19," "COVID 19"). The number of identified records were documented for each version of the term.

Proportion of MEDLINE-Indexed Records

To investigate the potential use of the indexing that is done when records are indexed in MEDLINE (ie, adding indexing terms such as MeSH and Supplementary Concepts), we calculated the proportion of records that had been MEDLINE-indexed out of the total number of records present in the LitCovid database by May 19, 2020. This was done by querying PubMed twice to retrieve the first 9999 PubMed IDs in the LitCovid database, followed by a second query to retrieve the remaining articles. We then looked at how many of the retrieved records out of the total number of retrieved records were tagged with the STAT – MEDLINE tag, indicating the status of the record as indexed in the MEDLINE database.

Software Analysis and Reproducibility Statement

All analyses were run in R (Version 4.0.2; R Foundation for Statistical Computing) and data were stored in text files when downloaded from PubMed using the PubMed format option. All search strings, code, and data to reproduce this analysis are available [27].

Results

Overview

In total, over the 10-week period studied, we found 13,599 records with our comprehensive search (Search 1). The Shokraneh search (Search 2; see Methods) found the largest number of records (n=13,880). The one-click search (Search 3) and the single-term search for COVID-19 (Search 4) yielded the same results, with the third-highest number of records (n=13,071). Next, the single-term search for Coronavirus (Search 6) found 9087 records, which was the fifth-highest

number, and the single-term search for SARS-CoV-2 (Search 5) found 7012 records, which was the sixth-highest number. The Wuhan coronavirus search (Search 7) found 5412 records and the Wuhan virus search (Search 8) found 1013 records. The number of records per week for each of the search strings is shown in Figure 1.

Figure 1. Records found over time from March 11 to May 19, 2020. Search 3, the one-click search, is not visible on the graph as it matches the results of the single-term search using COVID-19 (Search 4).
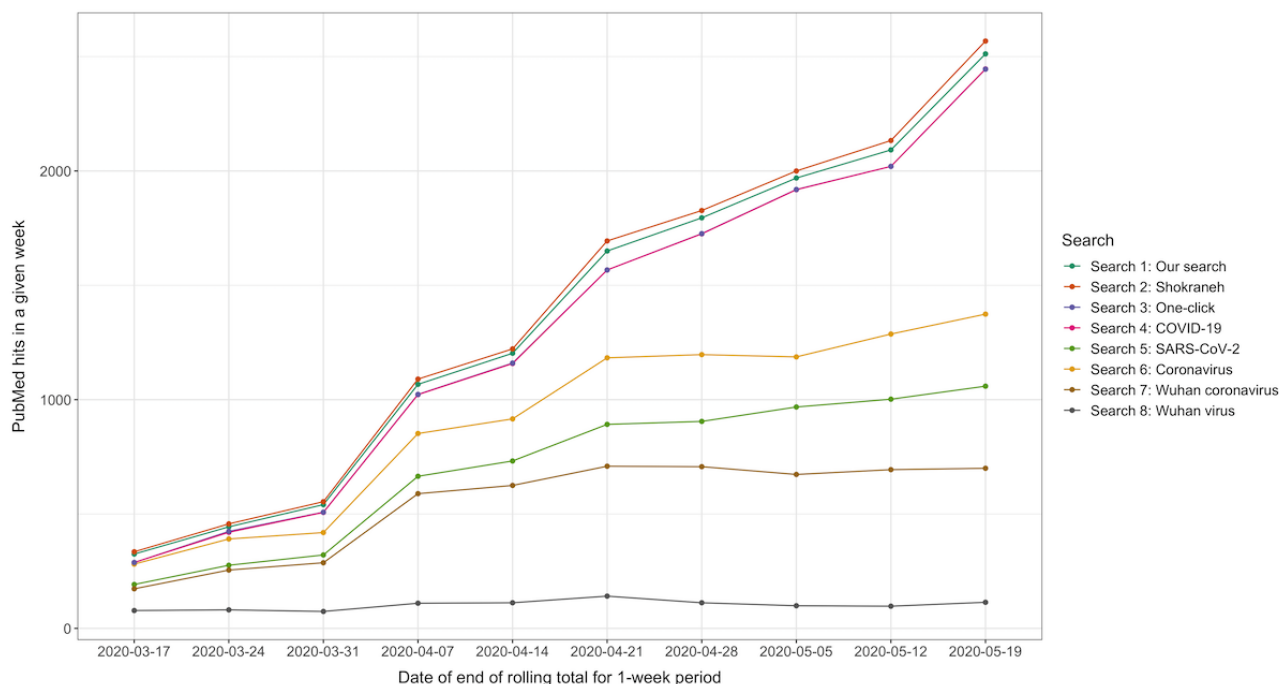
Evaluation of Search Strings

We found that the comprehensive searches (Searches 1 and 2) had the highest sensitivities and F-scores compared to all other searches. The searches for SARS-CoV-2 (Search 5) and Wuhan coronavirus (Search 7) had the highest precision (Table 3). The Wuhan virus search (Search 8) had the lowest sensitivity and precision, but otherwise the precision was similar for all the other searches. The difference in sensitivity of 4.0% between our comprehensive search (Search 1) and the single-term search for COVID-19 (Search 4) would translate to an average of 40 excess relevant records missed per 1000 articles identified when comparing the two searches over multiple theoretical sets of 1000 relevant COVID-19 records. This equates to an average of 43 excess relevant records missed if the single-term search COVID-19 (Search 4) was compared against Shokraneh's comprehensive search (Search 2) over multiple theoretical sets of 1000 relevant COVID-19 records.

Table 3. Metrics for the different strings as compared against the LitCovid gold standard.

	Records (n)	Sensitivity (%)	Precision (%)	F-score (%)
Search 1: Our comprehensive search	13,599	98.4	94.6	96.5
Search 2: Shokraneh's comprehensive search	13,880	98.7	92.7	95.7
Search 3: One-click search	13,071	94.4	95.3	94.8
Search 4: COVID-19	13,071	94.4	95.3	94.8
Search 5: SARS-CoV-2	7012	52.0	96.4	67.6
Search 6: Coronavirus	9087	67.2	93.4	78.3
Search 7: Wuhan coronavirus	5412	40.8	96.4	57.3
Search 8: Wuhan virus	1013	7.7	78.1	14.0

Sensitivity to Deleting Spaces and Hyphens and Adding Quotation Marks

We observed that automatic term mapping was sensitive to the deletion of hyphens and spaces, especially variations of SARS-CoV-2 (Table 4). We also found a decrease in records if a space or hyphen was removed from the search terms, such as COVID-19 versus COVID19 (13,071 versus 12,607). Furthermore, when quotation marks were added (eg, "COVID19"), the number of results fell to 11,716 records. Once again, SARS-CoV-2 terms were more sensitive to variations than COVID-19 terms.

The proportion of MEDLINE-indexed records out of the total number of records added to the LitCovid database during the

studied time period was 49.0% when verified in PubMed on July 5, 2020, for the PMIDs in the LitCovid data set from May 19, 2020.

Table 4. Analysis of searches with and without hyphens, spaces, and quotation marks.

Term	Result	Translation
covid-19	13,071	("COVID-19"[All Fields] OR "COVID-2019"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV-2"[All Fields] OR "2019nCoV"[All Fields] OR (("Wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields]))
covid 19	13,071	("COVID-19"[All Fields] OR "COVID-2019"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV-2"[All Fields] OR "2019nCoV"[All Fields] OR (("Wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields]))
covid19	12,607	"COVID-19"[Supplementary Concept] OR "COVID-19"[All Fields] OR "covid19"[All Fields]
"covid-19"	12,548	"covid-19"[All Fields]
"covid 19"	12,548	"covid 19"[All Fields]
"covid19"	11,716	"covid19"[All Fields]
sars-cov-2	7055	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "sars cov 2"[All Fields]
sars cov 2	7055	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "sars cov 2"[All Fields]
sars-cov 2	7055	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "sars cov 2"[All Fields]
sars cov-2	7055	"severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "sars cov 2"[All Fields]
sarscov-2	17	sarscov-2[All Fields]
sarscov 2	3008	sarscov[All Fields] AND 2[All Fields]
sars-cov2	153	sars-cov2[All Fields]
sars cov2	219	sars[All Fields] AND cov2[All Fields]
sarscov2	3601	sarscov2[All Fields]
"sars-cov-2"	3587	"sars-cov-2"[All Fields]
"sars cov 2"	3587	"sars cov 2"[All Fields]
"sars-cov 2"	3587	"sars-cov 2"[All Fields]
"sars cov-2"	3587	"sars-cov-2"[All Fields]
"sarscov-2"	17	"sarscov-2"[All Fields]
"sarscov 2"	17	"sarscov 2"[All Fields]
"sars-cov2"	153	"sars-cov2"[All Fields]
"sars cov2"	153	"sars cov2"[All Fields]
"sarscov2"	3601	"sarscov2"[All Fields]

Discussion

In this study, we evaluated eight PubMed searches and examined the differences between 24 alternative single-term searches with and without hyphens, spaces, and quotation marks. We found that the comprehensive search strings performed best in terms of sensitivity and F-score, while the one-click and single-term COVID-19 searches performed almost as well as the comprehensive search in terms of sensitivity and as well as the comprehensive search in terms of precision. The performance of the single-term COVID-19 search is dependent on PubMed's term mapping that translates the single-term search into a more comprehensive search. Comparatively, searching with SARS-CoV-2 as a single term while relying on the automatic term-mapping feature of PubMed yielded worse results than when searching with the single term COVID-19.

Using the LitCovid database, which covers both COVID-19 (the disease) and SARS-CoV-2 (the virus), as the gold standard comparator might have skewed our results in favor of the COVID-19 automatic term mapping. This is because the mapping translates the single-term search for COVID-19 to terms related to both the disease and the virus, whereas the single-term search for SARS-CoV-2 is translated to terms related

XSL•FO
RenderX

only to the virus and not the disease. This dual mapping does have precedent; PubMed's term mapping often maps the disease terms both to the virus and the disease, whereas the virus terms typically map only to the virus and not the disease.

The relatively early naming of the novel coronavirus, especially the name "COVID-19," set by the WHO, appears to have facilitated both the widespread use of the COVID-19 term in publications, as we found, and well-performing automatic term mapping in PubMed. As such, the results presented in this study highlight important progress in PubMed searching since the 2009 H1N1 influenza pandemic [17]. This progress could be further improved by NLM extending the mapping of the term COVID-19 to terms elucidated by evaluation of the two comprehensive searches in our analyses.

The analyses are based on our choices of the simple search terms we hypothesized users might enter. PubMed users might use other terms, such as Wuhan Pneumonia or COVID-2019 [28]. Nonetheless, the terms chosen by us consistently illustrate the differences between comprehensive and less comprehensive searches, whether constructed by users or via PubMed's automatic term mapping. Another limitation is the fact that MEDLINE indexing happens at different points in time after the record has been added to the PubMed database. This may result in different search results depending on the date of the search. Dates for all search results used in this study have been reported where appropriate. Finally, we have observed that the one-click search option has been changed since we conducted our analysis. Currently, the one-click search and the COVID-19 single-term search are identical. Thus, this has no implications on the interpretation of our result and conclusion.

Our sensitivity analyses of hyphens, spaces, and quotation marks also indicate room for improvement, especially when using SARS-CoV-2 as a single term for searching. As different writing style preferences and mistakes are unavoidable (eg, use of hyphens and spaces), automatic term mapping would be improved by being sensitive to this, just as it is sensitive to British and American spelling [23]. Surrounding the search term(s) with quotation marks forces an all-fields term or phrase search for the exact term(s) entered and does not activate automatic term mapping. This can markedly reduce the number of potentially relevant records in the search result as compared with a similar search based on a search string generated by automatic term mapping. The ability to turn off the automatic term mapping by adding quotation marks is not something that needs to be changed as it is a feature in PubMed. However, PubMed users need to be aware that adding quotation marks lowers sensitivity. In addition, the sensitivity analyses should motivate NLM to consider whether all variations (eg, COVID19) should activate the same automatic term mapping as COVID-19 (Search 4).

Implications for PubMed Users and NLM

Although some emphasize the importance of highly sensitive search strings more than others [15], it would be misleading to argue that the ability to identify all relevant articles on a given topic is relevant only for those conducting comprehensive, systematic reviews. Reviewers who conduct rapid reviews could save time and resources without substantially compromising

sensitivity and precision by using the search string from the one click-option. Everyday users of PubMed will need to specify the one click-search to reach the number of records they find manageable and relevant for their situation. However, the validity of the Best Match sorting option in PubMed rests on the sensitivity of the search process. Thus, the benefits of identifying all relevant records can extend to noncomprehensive PubMed searches.

Still, those who aim to conduct Cochrane-style systematic reviews would want to develop more comprehensive search strings rather than relying on the string integrated in the one click-option. Here, PubMed's Supplementary Concepts implemented for COVID-19 and SARS-CoV-2 could, if correct and consistently applied to all relevant records, help literature searchers conduct efficient searches. However, Supplementary Concepts are applied only to records available in PubMed that have been indexed in MEDLINE, which account for 49.0% of the total records identified in LitCovid. As reported above, the one-click PubMed search yielded the same results as the COVID-19 single-term search, although the latter included the Supplementary Concepts for both COVID-19 and SARS-CoV-2, suggesting that they do not add value when a search string is sufficiently comprehensive. For Supplementary Concepts to be of value for PubMed users aiming to conduct comprehensive reviews during the pandemic, NLM would have to speed up indexing of all records relevant to COVID-19.

We recommend that NLM uses a highly sensitive comprehensive search string to create a COVID-19 subject filter (ie, covid-19[sb]) or add it to their special queries collection [29,30]. The search string incorporated into such a filter or special query may even be activated by the automatic term mapping of a single-term search for COVID-19. If using one of the comprehensive search strings tested in this article to create the subject filter, it should be tested against other gold standard data sets for validation [31]. Future research should test more comprehensive search strings to determine which one is best suited for searching the literature base.

Further, we highlight the need for evaluating and validating search strings on multiple subjects (not only COVID-19) to develop more subject filters, which can be helpful for both everyday informational needs and serve as inspiration when conducting systematic reviews. Even so, we recommend that users consult with information specialists, research librarians, or researchers with the proper competencies for the retrieval of scientific information.

Conclusions

Scientific evidence must be easily accessible, especially during a pandemic. Overall, we found that changes have been made in PubMed that improve access to COVID-19–related articles compared to the situation during the 2009 H1N1 influenza pandemic. Importantly, some single-term searches performed well. Still, more can be done to support users searching for evidence regarding COVID-19. Specifically, the term mapping of the single-term COVID-19 search can be refined to be sensitive to variations in hyphens and spaces, and highly sensitive comprehensive search strings could be made more

easily available for instant application when using the PubMed search interface.

Overall, PubMed users can reliably use the one-click or single-term COVID-19 search for everyday informational needs about COVID-19 and SARS-CoV-2. However, when users are aiming to systematically locate and screen the total available literature on a topic related to COVID-19, especially when conducting systematic reviews, they should rely on comprehensive searches.

Authors' Contributions

JVL and ON conceived of the study. All authors contributed to the study design, data interpretation, and manuscript drafting.

Conflicts of Interest

JVL, AP, and JN have no competing interests. ON, THA, and LRN are employed at Steno Diabetes Center Copenhagen, a public hospital and research institution under the Capital Region of Denmark, which is partly funded by a grant from the Novo Nordisk Foundation. The funders had no role in this work.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nat Hum Behav 2020 Jul 23;4(7):666-669. [doi: 10.1038/s41562-020-0911-0] [Medline: 32576981]
3. Islamaj Dogan R, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford) 2009 Nov 27;2009:bap018-bap018 [FREE Full text] [doi: 10.1093/database/bap018] [Medline: 20157491]
4. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]
5. PubMed Overview 2020. National Library of Medicine. URL: https://pubmed.ncbi.nlm.nih.gov/about/ [accessed 2020-11-19]
6. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2007 Sep 20;22(2):338-342. [doi: 10.1096/fj.07-9492lsf]
7. Pubmed Subject Filter Strategies Updated for 2019. National Library of Medicine. URL: https://www.nlm.nih.gov/pubs/techbull/jf19/brief/jf19_pm_subject_filter_strategies_2019_update.html [accessed 2020-11-19]
8. National Library of Medicine. NCBI SARS-CoV-2 Resources. URL: https://www.ncbi.nlm.nih.gov/sars-cov-2 [accessed 2020-11-19]
9. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2019.
10. Bramer WM, Giustini D, Kleijnen J, Franco OH. Searching Embase and MEDLINE by using only major descriptors or title and abstract fields: a prospective exploratory study. Syst Rev 2018 Nov 20;7(1):200 [FREE Full text] [doi: 10.1186/s13643-018-0864-9] [Medline: 30458825]
11. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
12. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Med Res Methodol 2018 Aug 14;18(1):85 [FREE Full text] [doi: 10.1186/s12874-018-0545-3] [Medline: 30107788]
13. Booth A. Unpacking your literature search toolbox: on search styles and tactics. Health Info Libr J 2008 Dec;25(4):313-317 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00825.x] [Medline: 19076679]
14. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018 Oct 04;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]
15. Cooper C, Garside R, Varley-Campbell J, Talens-Bou J, Booth A, Britten N. "It has no meaning to me." How do researchers understand the effectiveness of literature searches? A qualitative analysis and preliminary typology of understandings. Res Synth Methods 2020 Sep;11(5):627-640. [doi: 10.1002/jrsm.1426] [Medline: 32495989]

XSL•FO

RenderX

16. Fiorini N, Canese K, Bryzgunov R, Radetska I, Gindulyte A, Latterner M, et al. PubMed Labs: an experimental system for improving biomedical literature search. Database (Oxford) 2018 Jan 01;2018:a [FREE Full text] [doi: 10.1093/database/bay094] [Medline: 30239682]

17. Norgaard O, Lazarus JV. Searching PubMed during a pandemic. PLoS One 2010 Apr 07;5(4):e10039 [FREE Full text] [doi: 10.1371/journal.pone.0010039] [Medline: 20383330]

18. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it [accessed 2020-11-19]

19. Daei A, Soleymani MR, Ashrafi-Rizi H, Zargham-Boroujeni A, Kelishadi R. Clinical information seeking behavior of physicians: A systematic review. Int J Med Inform 2020 Jul;139:104144. [doi: 10.1016/j.ijmedinf.2020.104144] [Medline: 32334400]

20. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature 2020 Mar 10;579(7798):193-193. [doi: 10.1038/d41586-020-00694-1] [Medline: 32157233]

21. Shokraneh F. Keeping up with studies on covid-19: systematic search strategies and resources. BMJ 2020 Apr 23;369:m1601. [doi: 10.1136/bmj.m1601] [Medline: 32327431]

22. Hodgkinson M. Pubmed PubReMiner. URL: https://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi [accessed 2020-11-19]

23. How PubMed works: Automatic Term Mapping. National Library of Medicine. URL: https://pubmed.ncbi.nlm.nih.gov/help/#automatic-term-mapping [accessed 2020-11-19]

24. Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. J Clin Epidemiol 2018 Jul;99:53-63. [doi: 10.1016/j.jclinepi.2018.02.025] [Medline: 29526555]

25. National Library of Medicine. PubMed User Guide - Clinical Queries. URL: https://pubmed.ncbi.nlm.nih.gov/help/#clinical-queries [accessed 2020-11-19]

26. Wilbur WJ, Kim W, Xie N. Spelling correction in the PubMed search engine. Inf Retr Boston 2006 Nov 1;9(5):543-564 [FREE Full text] [doi: 10.1007/s10791-006-9002-8] [Medline: 18080004]

27. Search strings, code, and data for this article. URL: https://osf.io/dae7j/ [accessed 2020-11-19]

28. Shokraneh F, Russell-Rose T. Lessons from COVID-19 to future evidence synthesis efforts: first living search strategy and out of date scientific publishing and indexing industry (submitted). J Clin Epidemiol 2020 Jul;123:171-173 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.04.014] [Medline: 32376119]

29. PubMed Subject Filters. National Library of Medicine. URL: https://www.nlm.nih.gov/bsd/pubmed_subsets.html [accessed 2020-11-19]

30. PubMed Special Queries. National Library of Medicine. URL: https://www.nlm.nih.gov/psd/special_queries.html [accessed 2020-11-19]

31. Tummers M, van Hoorn R, Levering C, Booth A, van der Wilt GJ, Kievit W. Optimal search strategies for identifying moderators and predictors of treatment effects in PubMed. Health Info Libr J 2019 Dec 13;36(4):318-340. [doi: 10.1111/hir.12230] [Medline: 30006959]

Abbreviations

MeSH: Medical Subject Headings
NLM: National Library of Medicine
WHO: World Health Organization

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Experiences of Psychotherapists With Remote Psychotherapy During the COVID-19 Pandemic: Cross-sectional Web-Based Survey Study

Elke Humer[1], PhD; Peter Stippl[2], PhD; Christoph Pieh[1], MD; Rüdiger Pryss[3], PhD; Thomas Probst[1], PhD

[1]Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

[2]Austrian Federal Association for Psychotherapy, Vienna, Austria

[3]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

Corresponding Author:
Thomas Probst, PhD
Department for Psychotherapy and Biopsychosocial Health
Danube University Krems
Dr-Karl-Dorrek-Straße 30
Krems, 3500
Austria
Phone: 43 27328932676
Email: thomas.probst@donau-uni.ac.at

Abstract

Background: The current situation around the COVID-19 pandemic and the measures necessary to fight it are creating challenges for psychotherapists, who usually treat patients face-to-face with personal contact. The pandemic is accelerating the use of remote psychotherapy (ie, psychotherapy provided via telephone or the internet). However, some psychotherapists have expressed reservations regarding remote psychotherapy. As psychotherapists are the individuals who determine the frequency of use of remote psychotherapy, the potential of enabling mental health care during the COVID-19 pandemic in line with the protective measures to fight COVID-19 can be realized only if psychotherapists are willing to use remote psychotherapy.

Objective: This study aimed to investigate the experiences of psychotherapists with remote psychotherapy in the first weeks of the COVID-19 lockdown in Austria (between March 24 and April 1, 2020).

Methods: Austrian psychotherapists were invited to take part in a web-based survey. The therapeutic orientations of the psychotherapists (behavioral, humanistic, psychodynamic, or systemic), their rating of the comparability of remote psychotherapy (web- or telephone-based) with face-to-face psychotherapy involving personal contact, and potential discrepancies between their actual experiences and previous expectations with remote psychotherapy were assessed. Data from 1162 psychotherapists practicing before and during the COVID-19 lockdown were analyzed.

Results: Psychotherapy conducted via telephone or the internet was reported to not be totally comparable to psychotherapy with personal contact ($P<.001$). Psychodynamic ($P=.001$) and humanistic ($P=.005$) therapists reported a higher comparability of telephone-based psychotherapy to in-person psychotherapy than behavioral therapists. Experiences with remote therapy (both web- and telephone-based) were more positive than previously expected ($P<.001$). Psychodynamic therapists reported more positive experiences with telephone-based psychotherapy than expected compared to behavioral ($P=.03$) and systemic ($P=.002$) therapists. In general, web-based psychotherapy was rated more positively (regarding comparability to psychotherapy with personal contact and experiences vs expectations) than telephone-based psychotherapy ($P<.001$); however, psychodynamic therapists reported their previous expectations to be equal to their actual experiences for both telephone- and web-based psychotherapy.

Conclusions: Psychotherapists found their experiences with remote psychotherapy (ie, web- or telephone-based psychotherapy) to be better than expected but found that this mode was not totally comparable to face-to-face psychotherapy with personal contact. Especially, behavioral therapists were found to rate telephone-based psychotherapy less favorably than therapists with other theoretical backgrounds.

XSL•FO
RenderX

Introduction

Background

Remote psychotherapy, in which psychotherapy is provided from a distance, includes a broad range of technologies, encompassing the use of telephones, videoconferencing, and email [1]. Mental health care conducted remotely (ie, via videoconferencing) has rapidly evolved worldwide as a technology, as it enables the direct delivery of real-time psychotherapy to patients [2]. The benefits of remote psychotherapy derive from the improved access to psychotherapy by providing mental health care services to patients who face logistical and stigma-related barriers to receiving face-to-face treatment [3].

The current situation around the COVID-19 pandemic and the measures necessary to fight it have further accelerated the rapid expansion of the technology of remote psychotherapy [4]. This is mainly because the traditional form of face-to-face psychotherapy conducted in person contrasts with the efforts to contain the COVID-19 pandemic, including social distancing, isolation, and quarantine [5]. However, the observed increase in mental health problems during the COVID-19 outbreak additionally enhances the general need for mental health care during and after the COVID-19 pandemic [6-8]. A recent review reported that the mental health problems related to quarantine include a high prevalence of psychological distress, depression, anxiety, and trauma-related disorders [6]. Thus, this public health emergency is enhancing the necessity to provide mental health care while adhering to efforts to contain the COVID-19 pandemic [9-11].

Due to the required reduction of personal contacts, psychotherapists are confronted with major challenges to delivery of care. In general, psychotherapists are the individuals who determine the frequency of use of remote psychotherapy. Thus, the potential of increasing access to mental health care during the COVID-19 pandemic while adhering to the protective measures to fight COVID-19 can be realized only if psychotherapists are willing to use remote psychotherapy. In this regard, it is important to investigate whether psychotherapists evaluate remote psychotherapy to be equal to face-to-face psychotherapy. Moreover, the potential of psychotherapists who have experience with remote psychotherapy to revise their attitudes based upon their recent experiences should be explored. A previous study suggested that health care providers who used telemedicine methods for the first time had more positive attitudes afterward [12]. Adequate mental health care is of high importance during the COVID-19 pandemic, and the provision of psychotherapy at a safe distance seems to be the obvious solution to ensure sufficient psychotherapeutic support. Thus, exploring the attitudes and experiences of psychotherapists toward remote psychotherapy is essential to improve the accessibility of mental health care systems during and after COVID-19.

Prior Work

In general, videoconferencing offers great potential for delivering psychotherapy from distance during the COVID-19 pandemic, as some evidence indicates comparable outcomes of providing psychotherapy remotely via the internet to in-person psychotherapy [13-15]. However, older technologies such as telephonic communication also offer immediate and easy-to-use ways to provide mental health care remotely. Moreover, equal effectiveness of telephone-based psychotherapy compared to face-to-face psychotherapy has been observed [16]. Nevertheless, despite research indicating comparable outcomes of providing psychotherapy remotely via telephone or the internet to providing in-person psychotherapy, psychotherapists have expressed some reservations about remote psychotherapy [17]. In general, psychotherapists seem to be more skeptical regarding remote psychotherapy than patients [2], who also show higher satisfaction with remote psychotherapy than therapists [18,19]. For example, while therapists rated the therapeutic relationship lower for videoconferencing than for in-person sessions, from the patients' perspective, no differences between these modalities were observed [20]. Moreover, technical difficulties seem to be experienced as more problematic by therapists than by patients [21].

Overall, several previous studies focused on the attitudes, experiences, and effectiveness of remote psychotherapy at the patient level; however, less is known about the therapists' perspective. In general, most studies comparing therapists' attitudes toward remote psychotherapy and face-to-face therapy found a preference for conducting therapy in person [2].

The acceptance and use of remote technologies by therapists may be affected by their theoretical background. Although there are hundreds of psychotherapeutic approaches and different ways to categorize them, they can be broadly categorized into four general schools of thought: behavioral, humanistic, psychodynamic, and systemic approaches. In brief, behavioral approaches mainly rely on behavioral techniques to change maladaptive patterns of behavior or thoughts to improve emotional responses and behaviors [22,23]. Humanistic or "experiential" psychotherapies are based on humanistic psychology, focusing mainly on human development and individual needs, with an emphasis on positive growth and subjective meaning [24]. Psychodynamic approaches focus on revealing or interpreting unconscious conflicts, which are thought to cause mental disorders [25]. In contrast, systemic therapy focuses on the interactions of groups such as families, as well as their dynamics and patterns, rather than addressing people individually. Systemic therapy seeks to identify and address stagnant patterns of behavior in groups [26,27]. Contrasting results have been obtained regarding the moderating role of these therapeutic orientations. Previous studies reported that psychodynamic orientation was related to more negative attitudes toward psychotherapy provided remotely [28-30], while behavioral orientations were found to be related to a more positive attitude toward remote psychotherapy [29,30]. Some research also indicates a higher acceptance of telehealth in

therapists with systemic orientations compared to psychodynamic or existential orientations [29]. However, other studies observed no relationship between therapeutic orientation and attitudes toward remote psychotherapy [31,32].

To the best of our knowledge, there are no previous studies on whether and to what degree psychotherapists perceive discrepancies in their actual experiences and previous expectations with remote psychotherapy in situations requiring a rapid adaption of therapeutic settings due to an ongoing public health emergency such as the COVID-19 outbreak. Therefore, the current study aimed to investigate how comparable psychotherapists experience remote psychotherapy compared to face-to-face psychotherapy and whether their actual experiences differ from their expectations. Furthermore, we were interested in potential differences among therapeutic orientations.

The COVID-19 lockdown in Austria became obligatory on March 16, 2020 [33-35]. In general, entering public places was strictly prohibited. People were only permitted to leave their homes if they had a good reason for doing so. At the time of the study, the only exceptions to the ban on entering public places were to avert immediate danger to life, limb, or property; to fulfill work responsibilities, although wherever possible, people should work from home; to meet necessary basic needs of daily life (eg, grocery shopping, visiting pharmacies, withdrawing money from cash machines, physician visits, medical treatments or therapy, pet maintenance); to take care of or support vulnerable people; and to practice low-risk sports (eg, walking or jogging), but only alone, with other people from one's own household, or with pets. For these exceptions, it was necessary to maintain a minimum safe distance of 1 m between people. Certain areas in Austria were under quarantine at the time of the study and had even stronger restrictions.

At the time of the study, an official Austrian guideline addressing the conduction of psychotherapy via the internet rejected this treatment modality [36]. However, health insurance started to cover the costs for telephone- and web-based psychotherapy during the COVID-19 pandemic in Austria.

Hypotheses

Based on the aforementioned literature, the following research questions (RQs) and hypotheses were addressed in the present study.

RQ 1: How do psychotherapists in the early weeks of the COVID-19 lockdown rate the comparability of telephone-based therapy to in-person psychotherapy?

RQ 1a: Do therapists rate telephone-based psychotherapy comparably to in-person psychotherapy? This RQ tested the hypothesis that psychotherapists would not rate telephone-based therapy to be equal to in-person therapy.

RQ 1b: Does the therapeutic orientation of the therapist affect the rating? We hypothesized that behavioral therapists would rate the comparability of telephone-based psychotherapy more positively compared to other therapeutic orientations. Moreover, we hypothesized that psychodynamic therapists would rate the comparability lower than other therapeutic orientations.

RQ 2: How much more negatively or positively do psychotherapists rate their actual experiences with telephone-based therapy in the early weeks of the COVID-19 lockdown compared to their previous expectations?

RQ 2a: Do the actual experiences of psychotherapists regarding telephone-based psychotherapy differ from their previous expectations? This RQ tested the hypothesis that therapists would rate their actual experiences with telephone-based psychotherapy higher than they previously expected.

RQ 2b: Is the discrepancy between actual experiences and previous expectations concerning telephone-based psychotherapy different between therapeutic orientations? We hypothesized that behavioral therapists would show a smaller discrepancy between their actual experiences and their previous expectations, whereas psychodynamic therapists would show the largest discrepancy.

RQ 3: How do psychotherapists in the early weeks of the COVID-19 lockdown rate the comparability of web-based therapy to in-person psychotherapy?

RQ 3a: Do therapists rate web-based psychotherapy comparably to in-person psychotherapy? This RQ tested the hypothesis that psychotherapists would not rate web-based therapy to be equal to in-person therapy.

RQ 3b: Does the therapeutic orientation of the psychotherapist affect the rating? We hypothesized that behavioral therapists would rate the comparability of web-based interventions more positively than therapists with other therapeutic orientations. Moreover, we hypothesized that psychodynamic therapists would rate the comparability lowest compared to the other therapeutic orientations.

RQ 4: How much more negatively or positively do psychotherapists rate their actual experiences with web-based therapy in the early weeks of the COVID-19 lockdown compared to their previous expectations?

RQ 4a: Do the actual experiences of psychotherapists regarding web-based psychotherapy differ from their previous expectations? This RQ tested the hypothesis that psychotherapists would rate their actual experiences with web-based therapy more positively compared to their previous expectations.

RQ 4b: Is the discrepancy between actual experiences and previous expectations different between therapeutic orientations? We hypothesized that behavioral therapists would show a smaller discrepancy between their actual experiences and their previous expectations, whereas psychodynamic therapists would show the largest discrepancy.

RQ 5: Does the format of remote psychotherapy (telephone or internet) affect the rating of psychotherapists regarding the comparability between remote psychotherapy and in-person psychotherapy? Moreover, is there an interaction between the comparability of telephone-based psychotherapy versus the comparability of web-based psychotherapy with the therapists' therapeutic orientations? We had no specific hypothesis here.

RQ 6: Does the discrepancy between actual experiences and previous expectations regarding telephone-based psychotherapy differ from the discrepancy between actual experiences and previous expectations regarding web-based psychotherapy? In addition, is there an interaction between actual experiences compared to previous expectations for telephone-based psychotherapy versus actual experiences compared to previous expectations for web-based psychotherapy and therapeutic orientation? We had no specific hypothesis here.

Methods

Study Design

To investigate the use of remote psychotherapy during the first weeks of the COVID-19 lockdown in Austria, we conducted a cross-sectional survey from March 24 to April 1, 2020. The survey was designed in the Research Electronic Data Capture (REDCap) application and comprised 79 items focusing on the changes in the provision of psychotherapy, experiences with remote psychotherapy, fear of COVID-19 infection, adherence to protective measures against COVID-19 infection, perceived stress, job-related anxiety, and resilience, among others. More details on the conduction of the web-based survey have been published recently [37,38].

Study Population

Eligible participants included all licensed Austrian psychotherapists registered in the list of psychotherapists of the Austrian Federal Ministry of Social Affairs, Health, Care, and Consumer Protection. All psychotherapists registered in the list who provided a valid email address (approximately 6000 out of 9319 licensed psychotherapists in total) were invited to take part in the survey. A web-based invitation with a link to the web-based survey was sent to the therapists by the last author in cooperation with the Austrian Federal Association for Psychotherapy.

Measures

The following items were analyzed in the current study: demographic information (gender, age) as well as items regarding educational and professional background. More specifically, the psychotherapists were asked about their years in the profession (ie, years since becoming accredited as a psychotherapist) and about their psychotherapy method (in Austria, there are 23 accredited therapeutic methods [39]); the latter was categorized in one of the four eligible therapeutic orientations in Austria (behavioral, humanistic, psychodynamic, and systemic) for the present study.

Four items focused on the expectations of the psychotherapists regarding telephone- and web-based therapies. These questions were only asked to psychotherapists who were treating at least one patient before the COVID-19 lockdown as well as after the COVID-19 lockdown.

Psychotherapists were asked to rate whether they could treat their patients in a comparable manner with remote therapy compared to in-person therapy on a sliding scale ranging from 0 (not comparable at all) to 100 (totally comparable). Two separate questions were asked, one on telephone-based psychotherapy, and the other on web-based psychotherapy. The questions on the comparability of telephone- or web-based psychotherapy were only asked of psychotherapists who treated at least one patient via telephone or via the internet either before the COVID-19 lockdown or during the COVID-19 lockdown.

Psychotherapists were asked to rate whether their actual experiences with remote psychotherapy during the actual situation around COVID-19 matched their previous expectations regarding remote psychotherapy on a sliding scale ranging from 0 (significantly more negative than expected) to 100 (significantly more positive than expected). Again, two separate questions were asked, focusing either on telephone- or web-based psychotherapy. The questions regarding actual experiences versus previous expectations for telephone- or web-based psychotherapy were only asked of psychotherapists who started to treat patients via telephone or the internet during the COVID-19 lockdown (ie, no patients before the COVID-19 lockdown via telephone or internet and at least one patient during the COVID-19 via telephone or the internet). Only "new starters" were asked these questions because they were the only respondents to show a change of attitude in a previous study focusing on health professionals in general [12].

Statistical Analysis

Statistical analyses were performed with SPSS version 25 (IBM Corporation).

Descriptive statistics were calculated to characterize the participants. To evaluate differences in sociodemographic characteristics, univariate analysis of variance (ANOVA) and chi-square tests were conducted. Bonferroni corrections were applied for the pairwise post-hoc tests.

The comparison of whether telephone- or web-based psychotherapy was regarded to be equal to psychotherapy in personal contact (RQ 1a, RQ 3a) was conducted using *t* tests; the respective ratings were compared with a value of 100, as a value of 100 represents the maximum score of the used sliding scale and means "totally comparable to psychotherapy in personal contact."

To evaluate possible discrepancies between actual experiences and previous expectations regarding the use of the telephone or internet for psychotherapy (RQ 2a, RQ 4a), the ratings were compared against a value of 50, as a value of 50 indicates "previous expectations=actual experiences."

To address RQ 1b, 2b, 3b, and 4b, univariate ANOVA was performed to investigate potential differences among the four therapeutic orientations of the psychotherapists (behavioral, humanistic, psychodynamic, and systemic) in the dependent variables (comparability of telephone- or web-based therapy to face-to-face therapy, actual experiences vs previous expectations for telephone- or web-based therapy). Bonferroni corrections were used for post-hoc comparisons.

To address RQ 5 and RQ 6, mixed ANOVA was performed with one within-subject variable (two levels of treatment format: telephone-based psychotherapy and web-based psychotherapy) and one between-subject variable (four levels of psychotherapeutic orientation: psychodynamic, humanistic,

systematic, behavioral). For variables with a significant interaction between treatment format and therapeutic orientation, Bonferroni corrections were applied for the pairwise post-hoc tests. The dependent variables were the comparability of telephone- or web-based therapy to face-to-face therapy and actual experiences versus previous expectations for telephone- or web-based therapy.

To evaluate whether the results are robust when accounting for the found differences in age and professional experience between the therapeutic orientations, we used analysis of covariance (ANCOVA) tests.

All statistical tests were two-tailed, with a P value <.05 indicating statistical significance.

Results

Participant Characteristics

In total, 1547 psychotherapists participated in the survey, corresponding to a response rate of approximately 25%. For the present study, only therapists practicing as psychotherapists in the months before the COVID-19 pandemic as well during the COVID-19 lockdown were analyzed to investigate the RQs. The description of the sample and comparisons between the included and excluded participants are shown in Table 1. There was a significant difference in age (P=.02); the excluded psychotherapists were older than the included psychotherapists. However, the effect size of this difference was small (g=0.14).

Table 1. Sociodemographic characteristics of the surveyed psychotherapists included in or excluded from the investigation (N=1547). Only respondents who practiced as psychotherapists before and during the COVID-19 pandemic were included.

Variable	Included participants (n=1162)	Excluded participants (n=385)	Statistical evaluation	P value
Gender			χ^2_1=2.454	.12
Female, n (%)	891 (76.7)	280 (72.7)		
Male, n (%)	271 (23.3)	105 (27.3)		
Age (years), mean (SD)	51.3 (9.67)	52.7 (9.69)	t_{1545}=2.409	.02
Therapeutic orientation[a]			χ^2_3=3.662	.30
Psychodynamic, n (%)	234 (20.3)	90 (23.7)		
Humanistic, n (%)	546 (47.4)	170 (44.9)		
Systemic, n (%)	252 (21.9)	88 (23.2)		
Behavioral, n (%)	120 (10.4)	31 (8.2)		
Professional experience (years),[b] mean (SD)	11.2 (9.21)	11.1 (9.19)	t_{1517}=–0.259	.80

[a]Data were available for 1152 of the included and 379 of the excluded participants.

[b]Data were available for 1143 of the included and 376 of the excluded participants.

For the 1162 included psychotherapists, we also investigated the differences between the therapeutic orientations in terms of gender, age, and professional experience (Table 2), with the following results:

Behavioral therapists were significantly younger than humanistic, systemic, and psychodynamic therapists (P≤.001 for all pairwise post-hoc comparisons). Similarly, behavioral therapists had significantly fewer years in the profession than humanistic, systemic, and psychodynamic therapists (P≤.04 for all pairwise post-hoc comparisons).

There was no difference regarding gender among the therapeutic orientations (P=.42).

XSL•FO

RenderX

Table 2. Sociodemographic characteristics of the psychotherapists included in the study in relation to their therapeutic orientation (n=1162).

Variable	n (%)	Mean years (SD)	Statistical evaluation	*P* value
Female gender			χ^2_3=2.797	.42
Behavioral	95 (79.2)	N/A[a]		
Humanistic	409 (74.9)	N/A		
Psychodynamic	187 (79.9)	N/A		
Systemic	195 (77.4)	N/A		
Age			$F_{3,1148}$=11.140	<.001
Behavioral	120 (10.4)	46.8 (9.82)		
Humanistic	546 (47.4)	51.7 (9.35)		
Psychodynamic	234 (20.3)	52.7 (9.96)		
Systemic	252 (21.9)	50.9 (9.37)		
Professional experience			$F_{3,1129}$=4.951	.002
Behavioral	117 (10.3)	8.4 (8.17)		
Humanistic	536 (47.3)	11.1 (9.23)		
Psychodynamic	229 (20.2)	12.4 (9.61)		
Systemic	251 (22.2)	11.2 (8.62)		

[a]N/A: not applicable.

Results for RQ 1

RQ 1a: Psychotherapists who treated at least one patient with telephone-based psychotherapy either before or during the COVID-19 lockdown (n=1015) stated that telephone-based psychotherapy is not totally comparable to in-person psychotherapy. The mean score of 50.2 (SD 26.00) was significantly lower than 100, T_{1014}=–61.07, *P*<.001.

RQ 1b: The therapeutic orientation affected the rating of whether telephone-based psychotherapy is comparable to face-to-face psychotherapy (*P*=.001). As summarized in Table 3, the highest values were observed for psychodynamic therapists, while the lowest values were observed for behavioral therapists. Post-hoc tests revealed a significantly higher rating by psychodynamic therapists compared to behavioral therapists (*P*=.001) and a higher rating by humanistic therapists compared to behavioral therapists (*P*=.005). No differences were observed for the systemic orientation compared to the other three orientations.

Table 3. Psychotherapists' ratings of their ability to treat patients via telephone comparable to in-person therapy on a sliding scale ranging from 0 (not comparable at all) to 100 (totally comparable) (n=1007; $F_{3,1003}$=5.626, *P*=.001).

Therapeutic orientation	n (%)	Mean (SD)
Behavioral	101 (10.0)	41.7 (24.97)
Humanistic	468 (46.5)	51.2 (25.64)
Psychodynamic	205 (20.4)	53.9 (25.78)
Systemic	233 (23.1)	48.6 (26.57)

Results for RQ 2

RQ 2a

Psychotherapists who started to use telephone for psychotherapy during the COVID-19 lockdown (n=782) stated that their actual experiences with telephone-based psychotherapy were better than previously expected. The mean score of 64.2 (SD 19.12) was significantly higher than 50 (t_{781}=20.707, *P*<.001).

RQ 2b

The therapeutic orientation affected the rating of whether telephone-based psychotherapy was experienced differently than expected (*P*=.001). As summarized in Table 4, the highest values were observed for psychodynamic therapists, while the lowest values were observed for behavioral therapists. Post-hoc tests revealed a significantly higher rating in psychodynamic compared to behavioral therapists, *P*=.03, and systemic therapists, *P*=.002. No differences were observed for the humanistic orientation compared to the other three orientations.

XSL•FO
RenderX

Table 4. Psychotherapists' actual experiences with telephone-based psychotherapy in relation to their previous expectations rated on a sliding scale ranging from 0 (significantly more negative than expected) to 100 (significantly more positive than expected) (n=777; $F_{3,773}$=5.243, P=.001).

Therapeutic orientation	n (%)	Mean (SD)
Behavioral	73 (9.4)	60.6 (21.87)
Humanistic	370 (47.6)	64.8 (17.48)
Psychodynamic	157 (20.2)	68.3 (18.98)
Systemic	177 (22.8)	60.8 (20.64)

Results for RQ 3

RQ 3a

Psychotherapists who treated at least one patient with web-based psychotherapy either before or during the COVID-19 lockdown (n=733) stated that web-based psychotherapy is not comparable to in-person psychotherapy. The mean score of 61.5 (SD 24.38) was significantly lower than 100 (t_{732}=–42.779, P<.001).

RQ 3b

The rating of whether web-based psychotherapy is comparable to face-to-face psychotherapy did not differ between therapeutic orientations (P=.88), as summarized in Table 5.

Table 5. Psychotherapists' rating of their ability to treat patients via the internet comparable to in-person therapy on a sliding scale ranging from 0 (not comparable at all) to 100 (totally comparable) (n=728; $F_{3,724}$=0.226, P=.88).

Therapeutic orientation	n (%)	Mean (SD)
Behavioral	85 (11.7)	60.1 (23.90)
Humanistic	352 (48.4)	61.6 (24.35)
Psychodynamic	128 (17.6)	61.3 (24.14)
Systemic	163 (22.4)	62.7 (24.74)

Results for RQ 4

RQ 4a

Psychotherapists who started to use internet for psychotherapy during the COVID-19 lockdown (n=614) stated that their actual experiences with web-based psychotherapy were better than previously expected. The mean score of 69.0 (SD 18.87) was significantly higher than 50 (t_{613}=24.88, P<.001).

RQ 4b

The discrepancy between the actual experiences and previous expectations regarding web-based psychotherapy did not differ concerning the therapeutic orientation of the psychotherapists (P=.71). The mean scores (SD) are summarized in Table 6.

Table 6. Psychotherapists' experiences with web-based psychotherapy in relation to their previous expectations on a sliding scale ranging from 0 (significantly more negative than expected) to 100 (significantly more positive than expected) (n=610; $F_{3,606}$=0.459, P=.71).

Therapeutic orientation	n (%)	Mean (SD)
Behavioral	72 (11.8)	67.7 (23.18)
Humanistic	303 (49.7)	68.7 (17.75)
Psychodynamic	100 (16.4)	68.7 (18.13)
Systemic	135 (22.1)	70.6 (19.39)

Results for RQ 5

Psychotherapists of the four orientations who treated at least one patient via telephone and at least one patient via the internet either before or during the COVID-19 lockdown (n=611) rated web-based psychotherapy to be more comparable to face-to-face psychotherapy than telephone-based psychotherapy ($F_{1,607}$=144.214, P<.001). No interaction between psychotherapy format (internet or telephone) and therapeutic orientation was observed ($F_{3,607}$=1.729, P=.16; Table 7).

Table 7. Psychotherapists' ratings of their ability to comparably treat their patients via telephone or the internet, respectively, compared to in-person therapy on a sliding scale ranging from 0 (not comparable at all) to 100 (totally comparable) (n=611; $F_{3,607}$=1.729, P=.16).

Therapeutic orientation	Telephone-based therapy		Web-based therapy	
	n (%)	Mean (SD)	n (%)	Mean (SD)
Behavioral	69 (11.3)	45.0 (24.25)	69 (11.3)	59.2 (24.51)
Humanistic	289 (47.3)	51.6 (24.41)	289 (47.3)	63.0 (23.41)
Psychodynamic	105 (17.2)	52.6 (24.74)	105 (17.2)	60.3 (24.36)
Systemic	148 (24.2)	50.1 (25.49)	148 (24.2)	62.8 (23.92)

Results for RQ 6

Psychotherapists of the four therapeutic orientations who started to treat at least one patient via telephone and at least one patient via the internet during the COVID-19 lockdown (n=422) reported that their actual experiences with web-based psychotherapy were more positive than their actual experiences with telephone-based psychotherapy compared to their previous expectations ($F_{1,418}$=22.680; P<.001). An interaction between the psychotherapy format (internet or telephone) and the

therapeutic orientation was observed ($F_{3,418}$=4.862, P=.002; Table 8). Post-hoc comparisons revealed that behavioral (P=.03), humanistic (P=.005), and systemic (P<.001) psychotherapists reported that their actual experiences compared to their previous expectations were more positive for web-based therapy than for telephone-based therapy. However, the discrepancy between actual experiences and previous expectations with remote psychotherapy was not different between web- and telephone-based psychotherapy among psychodynamic therapists (P>.99).

Table 8. Psychotherapists' experiences with telephone- and web-based psychotherapy, respectively, compared to previous expectations rated on a sliding scale ranging from 0 (significantly more negative than expected) to 100 (significantly more positive than expected) (n=422; $F_{3,418}$=4.862, P=.002).

Therapeutic orientation	Telephone-based therapy		Web-based therapy	
	n (%)	Mean (SD)	n (%)	Mean (SD)
Behavioral	45 (10.7)	63.2 (22.21)	45 (10.7)	69.0 (22.05)
Humanistic	202 (47.9)	66.1 (16.91)	202 (47.9)	69.7 (17.58)
Psychodynamic	72 (17.1)	68.3 (16.58)	72 (17.1)	68.3 (18.22)
Systemic	103 (24.4)	61.7 (21.13)	103 (24.4)	71.6 (19.95)

Discussion

Principal Results

This survey explored the experiences of psychotherapists in Austria with remote psychotherapy during the COVID-19 lockdown. Remote psychotherapy was reported to not be totally comparable to in-person psychotherapy, although the psychotherapists reported that their actual experiences with remote psychotherapy were better than expected. Psychodynamic and humanistic therapists rated telephone-based psychotherapy more comparably to face-to-face psychotherapy than behavioral therapists. The actual experiences with telephone-based psychotherapy differed more positively from previous expectations among psychodynamic therapists than among behavioral and systemic therapists. Web-based psychotherapy was rated to be more comparable to in-person psychotherapy than telephone-based psychotherapy, and this did not differ between therapeutic orientations. Also, the actual experiences of most psychotherapists were more positive than their previous expectations for web-based psychotherapy than for telephone-based psychotherapy; however, this was not the case for psychodynamic therapists.

Limitations

There are several limitations in this study. One limitation is that the survey was conducted on the web; this may have caused some respondent bias, such as higher participation of psychotherapists with a higher preference for new technologies. This bias may have contributed to the finding that providing psychotherapy remotely was rated more positively than expected. The web-based conduction of the survey may also have caused a selection bias toward participation by fewer older psychotherapists [40]. A further limitation is that "telephone-based" and "web-based" are rather broad categories for treatment formats, and more detailed information about these formats, such as use of videoconferencing, chats, apps, or email, was not assessed. Another shortcoming is that the comparability of remote psychotherapy to in-person psychotherapy was operationalized by a survey conducted among psychotherapists, while no patient rating surveys or effectiveness measures were conducted. For future studies, the patients' perspective, as well as outcome measures, should also be evaluated. These studies are not easily performed with web-based surveys; randomized controlled trials to evaluate efficacy need considerable planning time and thus were not feasible to evaluate remote psychotherapy during the first weeks of the COVID-19 lockdown. A further shortcoming is that the web-based conduction of the study prevented any measures of treatment

adherence. Thus, it is not possible to say whether the therapeutic methods applied truly resembled the theoretical methods the therapists were trained to use. Moreover, the included participants are not representative of the excluded survey participants in age, as the excluded participants were approximately 1.5 years older (the effect size of this difference was rather small, with g=0.14). As the study was conducted in Austria, the results may only apply to countries with similar mental health care systems and similar therapeutic orientations among psychotherapists. In Austria, internet-based psychotherapy was rejected by an official guideline at the time of the study [34]. However, health insurance companies started to cover the costs of telephone- and web-based psychotherapy during the COVID-19 pandemic. Thus, the results are not directly comparable to countries that had already implemented mental eHealth solutions in routine psychotherapeutic practice.

Comparison With Prior Work

Our results confirm the hypothesis that psychotherapy via telephone or the internet is not regarded to be completely comparable to in-person psychotherapy by therapists. In agreement, most studies investigating experience with remote psychotherapy reported that a minority of therapists experienced remote psychotherapy to be equal to face-to-face encounters [20,21,41]. Doubts regarding the comparability of remote psychotherapy with face-to-face psychotherapy arise from low performance expectancy, the lack of nonverbal communication, and difficulties in dealing with crises from a distance [17]. Also, the unsuitability of remote psychotherapy for all patients was identified as a disadvantage, and some therapists were concerned that remote therapy would be time-consuming or hinder the establishment of a therapeutic relationship [42].

Our results further confirm the hypothesis that therapists who started to use remote psychotherapy had better experiences with remote psychotherapy than previously expected; this was more pronounced for web-based psychotherapy than for telephone-based psychotherapy. This result is supported by previous studies showing positive experiences of therapists using remote psychotherapy [43-45]. As reviewed by Connolly et al [2], the functionality and ease of use of psychotherapy provided via videoconferencing was reported to be a "pleasant surprise." The generally positive experiences with remote psychotherapy can be explained by benefits reported previously by therapists, such as improved access to therapy and less traveling time [41]. However, the special situation of the COVID-19 pandemic may have further benefitted the favorable rating of remote psychotherapy, as it represented the only way to maintain psychotherapy while adhering to the protective measures against COVID-19. In general, psychotherapists are the individuals who determine the frequency of use of remote psychotherapy. The finding that psychotherapists experience remote psychotherapy more positively than previously expected suggests that it is important to provide a practical experience (already in training); this finding is also in line with a previous study on changing attitudes of health care providers after first-time use of telemedicine [12].

The findings were also influenced by the participants' therapeutic orientation. In contrast to previous findings, behavioral therapists reported less positive actual experiences compared to previous expectations and lower comparability of telephone-based psychotherapy to in-person therapy than psychodynamic therapists [28-30]. This result contrasts with our hypothesis, which assumed that psychodynamic psychotherapists would rate telephone-based psychotherapy more negatively than behavioral psychotherapists. Previously, it was assumed that psychodynamic therapists would particularly feel that they were losing valuable information through remote psychotherapy, even when using videoconferencing. The analysis of nonverbal behavior is of paramount importance in psychodynamic theory. Thus, information regarding behavioral or physiological changes (such as crossing or swinging a foot, muscle tension, or perspiration) is difficult to perceive in remote settings, even when using a camera [29]. However, our results do not support the assumption that this loss of information will have a detrimental effect on the attitudes and experiences of psychodynamic therapists toward remote psychotherapy. In contrast, psychotherapy via telephone, which causes a much stronger loss of information related to nonverbal behavior than videoconferencing, was rated more positively (regarding comparability to face-to-face therapy and experiences vs expectations) by psychodynamic therapists than by therapists of other therapeutic orientations. However, ratings of comparability to face-to-face psychotherapy were also higher for web-based therapy than for telephone-based therapy for psychodynamic therapists. On the other hand, the lowest ratings for telephone-based therapy (regarding comparability to face-to-face therapy and experiences vs expectations) were found for behavioral therapists compared to therapists of other therapeutic orientations. Because behavioral therapy is more focused on changing maladaptive behaviors [29], we assumed that those therapists would rate remote psychotherapy more positively than other therapists, as these maladaptive behaviors could be treated similarly in remote and face-to-face therapy. However, it is possible that the observed differences among therapeutic orientations are confounded by other factors, such as differences in gender, age, or therapeutic experiences, as those factors have been speculated to be potential moderators of attitudes toward remote psychotherapy [2,46]. To rule out any confounding effect of the observed differences concerning age and experience (ie, the younger age and shorter period in the profession of behavioral therapists), those variables were included as covariates in ANCOVAs. However, even when conducting ANCOVAs, differences among therapeutic orientations remained significant. Thus, further studies are required to investigate the potential reasons behind the observed differences.

The ratings of the comparability of web-based and in-person psychotherapy as well as the discrepancies in actual experiences and previous expectations regarding web-based psychotherapy did not differ regarding therapeutic background. This result rejects our hypothesis that behavioral psychotherapists would rate remote psychotherapy via internet more positively than psychodynamic therapists.

In general, web-based psychotherapy was rated more positively (regarding comparability to face-to-face therapy and experiences vs expectations) than psychotherapy via telephone; however,

this was not the case for psychodynamic therapists, who reported their previous expectations to be equal to their actual experiences for both telephone- and web-based psychotherapy. This confirms our hypothesis and is likely due to the potential of videoconferencing to provide several pieces of information regarding nonverbal behavior and of its greater comparability to face-to-face settings than telephone-based psychotherapy, which results in a greater loss of information.

Conclusions

The experiences of psychotherapists with remote psychotherapy were better than their expectations but not totally comparable

to face-to-face psychotherapy with personal contact. Adequate mental health care is of high importance during the COVID-19 pandemic, and the provision of psychotherapy at a safe distance seems to be the obvious solution to ensure sufficient psychotherapeutic support. This study provides important insights regarding psychotherapists' experiences with and expectations of remote psychotherapy during the COVID-19 pandemic. Future studies should also consider patient perspectives.

Conflicts of Interest

None declared.

References

1. Grady B, Myers K, Nelson EL, Belz NLL, Bennett L, Carnahan L, American Telemedicine Association Telemental Health Standards and Guidelines Working Group. Evidence-based practice for telemental health. Telemed J E Health 2011 Mar;17(2):131-148. [doi: 10.1089/tmj.2010.0158] [Medline: 21385026]

2. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol Sci Pract 2020 Jan 06;27(2):e12311. [doi: 10.1111/cpsp.12311]

3. Fletcher TL, Hogan JB, Keegan F, Davis ML, Wassef M, Day S, et al. Recent Advances in Delivering Mental Health Treatment via Video to Home. Curr Psychiatry Rep 2018 Jul 21;20(8):56. [doi: 10.1007/s11920-018-0922-y] [Medline: 30032337]

4. Wright J, Caudill R. Remote Treatment Delivery in Response to the COVID-19 Pandemic. Psychother Psychosom 2020 Mar 26;89(3):130-132 [FREE Full text] [doi: 10.1159/000507376] [Medline: 32213775]

5. Nussbaumer-Streit B, Mayr V, Dobrescu AJ, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database Syst Rev 2020 Apr 08;4:CD013574. [doi: 10.1002/14651858.CD013574] [Medline: 32267544]

6. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]

7. Hossain M, Sultana A, Purohit N. Mental health outcomes of quarantine and isolation for infection prevention: A systematic umbrella review of the global evidence. Research Square. Preprint posted online on April 28, 2020. [doi: 10.21203/rs.3.rs-25647/v1]

8. Sharma A, Pillai D, Lu M, Doolan C, Leal J, Kim J, et al. Impact of isolation precautions on quality of life: a meta-analysis. J Hosp Infect 2020 May;105(1):35-42. [doi: 10.1016/j.jhin.2020.02.004] [Medline: 32059996]

9. Fiorillo A, Gorwood P. The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. Eur Psychiatry 2020 Apr 01;63(1):e32 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.35] [Medline: 32234102]

10. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]

11. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiat 2020 Mar;7(3):228-229. [doi: 10.1016/s2215-0366(20)30046-8]

12. Hanson D, Calhoun J, Smith D. Changes in provider attitudes toward telemedicine. Telemed J E Health 2009 Jan;15(1):39-43. [doi: 10.1089/tmj.2008.0052] [Medline: 19199846]

13. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The Empirical Evidence for Telemedicine Interventions in Mental Disorders. Telemed J E Health 2016 Feb;22(2):87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]

14. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]

15. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]

16. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 06;307(21):2278-2285 [FREE Full text] [doi: 10.1001/jama.2012.5588] [Medline: 22706833]

17.   Schuster R, Pokorny R, Berger T, Topooco N, Laireiter A. The Advantages and Disadvantages of Online and Blended Therapy: Survey Study Amongst Licensed Psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]

18.   Shulman M, John M, Kane JM. Home-Based Outpatient Telepsychiatry to Improve Adherence With Treatment Appointments: A Pilot Study. Psychiatr Serv 2017 Jul 01;68(7):743-746 [FREE Full text] [doi: 10.1176/appi.ps.201600244] [Medline: 28245700]

19.   Thomas JF, Novins DK, Hosokawa PW, Olson CA, Hunter D, Brent AS, et al. The Use of Telepsychiatry to Provide Cost-Efficient Care During Pediatric Mental Health Emergencies. Psychiatr Serv 2018 Feb 01;69(2):161-168. [doi: 10.1176/appi.ps.201700140] [Medline: 29032703]

20.   Ertelt TW, Crosby RD, Marino JM, Mitchell JE, Lancaster K, Crow SJ. Therapeutic factors affecting the cognitive behavioral treatment of bulimia nervosa via telemedicine versus face-to-face delivery. Int J Eat Disord 2011 Dec 15;44(8):687-691 [FREE Full text] [doi: 10.1002/eat.20874] [Medline: 22072405]

21.   Schopp L, Johnstone B, Merrell D. Telehealth and neuropsychological assessment: New opportunities for psychologists. Prof Psychol Res Pr 2000 Apr;31(2):179-183. [doi: 10.1037/0735-7028.31.2.179]

22.   DiTomasso A, Golden B, Morris H. Handbook of Cognitive Behavioral Approaches in Primary Care. New York, NY: Springer; 2009.

23.   Sundberg N, Winebarger A, Taplin J. Clinical Psychology: Evolving Theory, Practice, and Research (4th Edition). London, UK: Pearson; 2001.

24.   Sajikumar D. In: Sandeep VS, editor. New holistic psychotherapies. Kothamangalam, India: Dr V Sajikumar; 2020.

25.   Cabaniss D, Cherry S, Douglas C, Schwartz A. Psychodynamic Psychotherapy: A Clinical Manual (2nd Edition). Hoboken, NJ: Wiley-Blackwell; 2016.

26.   Pocock D. A philosophy of practice for systemic psychotherapy: the case for critical realism. J Fam Ther 2013 Dec 10;37(2):167-183. [doi: 10.1111/1467-6427.12027]

27.   Schlippe A, Schweitzer J. Lehrbuch der systemischen Therapie und Beratung II. Das störungsspezifische Wissen (3rd Edition). Göttingen, Germany: Vadenhoeck & Ruprecht; 2015.

28.   Gordon RM, Wang X, Tune J. Comparing Psychodynamic Teaching, Supervision, and Psychotherapy Over Videoconferencing Technology with Chinese Students. Psychodyn Psychiatry 2015 Dec;43(4):585-599. [doi: 10.1521/pdps.2015.43.4.585] [Medline: 26583442]

29.   Perle JG, Langsam LC, Randel A, Lutchman S, Levine AB, Odland AP, et al. Attitudes toward psychological telehealth: current and future clinical psychologists' opinions of internet-based interventions. J Clin Psychol 2013 Jan;69(1):100-113. [doi: 10.1002/jclp.21912] [Medline: 22975897]

30.   Wangberg SC, Gammon D, Spitznogle K. In the eyes of the beholder: exploring psychologists' attitudes towards and use of e-therapy in Norway. Cyberpsychol Behav 2007 Jun;10(3):418-423. [doi: 10.1089/cpb.2006.9937] [Medline: 17594266]

31.   Cipolletta S, Mocellin D. Online counseling: An exploratory survey of Italian psychologists' attitudes towards new ways of interaction. Psychother Res 2018 Nov 09;28(6):909-924. [doi: 10.1080/10503307.2016.1259533] [Medline: 28068875]

32.   Eichenberg C, Kienzle K. Psychotherapeuten und Internet. Psychotherapeut 2011 Jun 11;58(5):485-493. [doi: 10.1007/s00278-011-0833-4]

33.   98. Verordnung: Verordnung gemäß § 2 Z 1 des COVID-19-Maßnahmengesetzes. Bundesgesetzblatt für die Republik Österreich. 2020. URL: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2020_II_98/BGBLA_2020_II_98.html [accessed 2020-11-24]

34.   107. Verordnung: Änderung der Verordnung gemäß § 2 Z 1 des COVID-19-Maßnahmengesetzes. Bundesgesetzblatt für die Republik Österreich. 2020. URL: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2020_II_107/BGBLA_2020_II_107.html [accessed 2020-05-11]

35.   108. Verordnung: Änderung der Verordnung gemäß § 2 Z 1 des COVID-19-Maßnahmengesetzes. Bundesgesetzblatt für die Republik Österreich. 2020. URL: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2020_II_108/BGBLA_2020_II_108.html [accessed 2020-05-11]

36.   Internetrichtlinie Kriterien zur Ausgestaltung der psychotherapeutischen Beratung via Internet. Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz. URL: https://www.sozialministerium.at/Themen/Gesundheit/Medizin-und-Gesundheitsberufe/Erlässe,-Empfehlungen-und-Informationen.html [accessed 2020-05-11]

37.   Probst T, Stippl P, Pieh C. Changes in Provision of Psychotherapy in the Early Weeks of the COVID-19 Lockdown in Austria. Int J Environ Res Public Health 2020 May 27;17(11):3815 [FREE Full text] [doi: 10.3390/ijerph17113815] [Medline: 32471295]

38.   Probst T, Humer E, Stippl P, Pieh C. Being a Psychotherapist in Times of the Novel Coronavirus Disease: Stress-Level, Job Anxiety, and Fear of Coronavirus Disease Infection in More Than 1,500 Psychotherapists in Austria. Front Psychol 2020 Sep 29;11:559100. [doi: 10.3389/fpsyg.2020.559100] [Medline: 33132965]

39.   Heidegger KE. The Situation of Psychotherapy in Austria. Österreichischer Bundesverband für Psychotherapie. 2017. URL: https://www.europsyche.org/app/uploads/2019/05/Situation-Psychotherapy-in-Austria-2017-10-20.pdf [accessed 2020-07-14]

40.   Bethlehem J. Selection bias in web surveys. Int Stat Rev 2010;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]

41. Glover JA, Williams E, Hazlett LJ, Campbell N. Connecting to the future: telepsychiatry in postgraduate medical education. Telemed J E Health 2013 Jun;19(6):474-479. [doi: 10.1089/tmj.2012.0182] [Medline: 23570291]

42. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJEWC. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]

43. Adler G, Pritchett LR, Kauth MR, Nadorff D. A pilot project to improve access to telepsychotherapy at rural clinics. Telemed J E Health 2014 Jan;20(1):83-85 [FREE Full text] [doi: 10.1089/tmj.2013.0085] [Medline: 24168721]

44. Elford R, White H, Bowering R, Ghandi A, Maddiggan B, St John K, et al. A randomized, controlled trial of child psychiatric assessments conducted using videoconferencing. J Telemed Telecare 2000 Jun 24;6(2):73-82. [doi: 10.1258/1357633001935086] [Medline: 10824374]

45. Lindsay JA, Hudson S, Martin L, Hogan JB, Nessim M, Graves L, et al. Implementing Video to Home to Increase Access to Evidence-Based Psychotherapy for Rural Veterans. J Technol Behav Sci 2017 Dec 28;2(3-4):140-148 [FREE Full text] [doi: 10.1007/s41347-017-0032-4] [Medline: 32346604]

46. Simms DC, Gibson K, O'Donnell S. To use or not to use: Clinicians' perceptions of telemental health. Can Psychol 2011 Feb;52(1):41-51. [doi: 10.1037/a0022275]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
REDCap: Research Electronic Data Capture
RQ: research question

XSL•FO
RenderX

Original Paper

Novel Indicator to Ascertain the Status and Trend of COVID-19 Spread: Modeling Study

Takashi Nakano[1*], DPhil; Yoichi Ikeda[2*], DPhil

[1]Research Center for Nuclear Physics, Osaka University, Osaka, Japan

[2]Department of Physics, Faculty of Science, Kyushu University, Fukuoka, Japan

[*]all authors contributed equally

Corresponding Author:
Takashi Nakano, DPhil
Research Center for Nuclear Physics
Osaka University
10-1 Mihogaoka
Ibaraki
Osaka, 567-0047
Japan
Phone: 81 6 6879 8900
Fax: 81 6 6879 8899
Email: nakano@rcnp.osaka-u.ac.jp

Abstract

Background: In the fight against the pandemic of COVID-19, it is important to ascertain the status and trend of the infection spread quickly and accurately.

Objective: The purpose of our study is to formulate a new and simple indicator that represents the COVID-19 spread rate by using publicly available data.

Methods: The new indicator K is a backward difference approximation of the logarithmic derivative of the cumulative number of cases with a time interval of 7 days. It is calculated as a ratio of the number of newly confirmed cases in a week to the total number of cases.

Results: The analysis of the current status of COVID-19 spreading over countries showed an approximate linear decrease in the time evolution of the K value. The slope of the linear decrease differed from country to country. In addition, it was steeper for East and Southeast Asian countries than for European countries. The regional difference in the slope seems to reflect both social and immunological circumstances for each country.

Conclusions: The approximate linear decrease of the K value indicates that the COVID-19 spread does not grow exponentially but starts to attenuate from the early stage. The K trajectory in a wide range was successfully reproduced by a phenomenological model with the constant attenuation assumption, indicating that the total number of the infected people follows the Gompertz curve. Focusing on the change in the value of K will help to improve and refine epidemiological models of COVID-19.

Introduction

The spread of COVID-19 has resulted in human and economic losses worldwide. To prevent the spread of the infection, it is sometimes necessary to restrict social activities by policies such as the blockade of cities and the prohibition of assembly. For the effective implementation of these policies, it is important to ascertain the status and trend of spread quickly and accurately.

The purpose of our study is to formulate a simple indicator that represents the COVID-19 spread rate by using publicly available data [1-4] and to reveal the common features and regional differences of the spread. We also aim to check the validity of the widely accepted assumption that the number of daily confirmed cases can be approximated by the exponential function in the early stage of COVID-19 spread.

XSL•FO
RenderX

Methods

To analyze the trend of COVID-19 spread, we introduced a new indicator K as a measure of the spread rate. It is defined by $K(t) = 1 - N(t - 7)/N(t)$, where t is the number of days from a reference date and $N(t)$ and $N(t - 7)$ are the total number of cases on days t and $(t - 7)$, respectively. Thus, the K value is a backward difference approximation of the logarithmic derivative of $N(t)$ with a time interval of 7 days. If $N(t)$ grows exponentially as obtained in the susceptible-infected-removed (SIR) model [5-10] during the early stage of the epidemic, K takes a constant value. For example, if $N(t)$ keeps doubling in 7 days from the reference date, the corresponding K value is 0.5. Since $N(t)$ is greater than $N(t - 7)$ during the period from the initiation of spread to convergence, the range of the K value is between 0 and 1. We note that we can cancel the day of the week dependency seen in the number of daily confirmed cases by setting the interval to 7 days.

Results

The K values for China were calculated by setting the reference date to January 26, 2020, where the data before February 12 were uniformly multiplied by 1.27 to correct a discontinuous increase in the number of infected people caused by the change of certification criteria for SARS-CoV-2 infections in Hubei Province on February 13. As shown in Figure 1, the trajectory of the K values is well approximated by a straight line with a slope (K') of $-0.0402 \pm 0.0008/d$. This linearly decreasing behavior evidently shows that the COVID-19 spread does not follow the exponential growth even in the early stage [8] but most likely follows a double exponential function known as the Gompertz curve (see the proceeding section, Interpretation of the K transition from a phenomenological model). The linearity of K is also prominent in the United States. After a high K level period indicating successive infectious explosions from mid- to late March, the K continued to decline with a uniform rate of $K'=-0.0237 \pm 0.0003/d$.

Countries' policies affected the transitions of the K values in Europe. In Italy, where COVID-19 started to spread first in Europe, K' was $-0.0142 \pm 0.0004/d$ from March 1 to March 23, 2020, indicating a slow pace of convergence, but the pace was improved after the implementation of a containment policy on March 24, resulting in $K'=-0.0263 \pm 0.0006/d$. In France, where the spread started about 10 days after Italy, the slope K' was $-0.0152 \pm 0.0005/d$ for 3 weeks in the early stage, but it was improved to $-0.0265 \pm 0.0019/d$ after the lockdown. Germany began measures to prevent the spread of infection, such as closure of most retail stores on March 16, 2020, resulting in a steep slope ($K'=-0.0252 \pm 0.0005/d$) afterward. Sweden, on the other hand, sought to acquire herd immunity and took relatively mild measures, resulting in a moderate slope ($K'=-0.0179 \pm 0.0004/d$) lasting more than a month. The United Kingdom took similar measures in the early stage, resulting in infection explosion with $K'=-0.0071 \pm 0.0008/d$. However, the slope became $-0.0199 \pm 0.0005/d$ after the introduction of stricter policies. In Russia, the K value stayed above 0.6 at $K'=-0.0067 \pm 0.0005/d$ until April 20 and then started to decrease at $K'=-0.0234 \pm 0.0010/d$, indicating that a catastrophic situation was avoided.

In Asian countries, after the first wave originated from China, the subsequent spread worldwide caused the upward change in K trajectories. To measure the rate of the subsequent spread in terms of the K value accurately, the reference date had to be set at the rise of the spread. Thus, the K values for Japan were calculated by setting the reference date to March 25, 2020. The obtained slope ($K'=-0.0283 \pm 0.0006/d$) was milder than those of Taiwan ($K'=-0.0524 \pm 0.0026/d$) and South Korea ($K'=-0.0820 \pm 0.0042/d$ and $K'=-0.0378 \pm 0.0024/d$), reflecting the difference in the strictness of countermeasures. However, it was steeper than those of European countries with more strict social restrictions than Japan. The relatively high absolute K' values even in the early stages were commonly observed in many Asian countries. For example, the K' value of Thailand is $K'=-0.0361 \pm 0.0028/d$.

XSL·FO
RenderX

Figure 1. Transition of the K values from February to April 2020. (A) The K values of China obtained from the daily total number of infected people. The slope K' was obtained by a linear fit in the range of $0.25 < K < 0.9$. The data points used for the fit are indicated by red points. The solid line is the fit result. (B) The K values of the United States. (C) The K values of Italy. The first and second K' values were obtained by linear fits using the red and green points, respectively. The solid line is the result of the first fit. The dashed line is the result of the second fit. (D) The K values of France. (E) The K values of Germany. (F) The K values of Sweden. The total number of infected people was counted from the reference date set for March 14. (G) The K values of the United Kingdom. (H) The K values of Russia. (I) The K values of Japan. The reference date was set for March 25. (J) The K values of Taiwan. The reference date was set for March 12. (K) The K values of South Korea. The reference dates were set for February 25 and March 18 for the first and second fits, respectively. (L) The K values of Thailand.
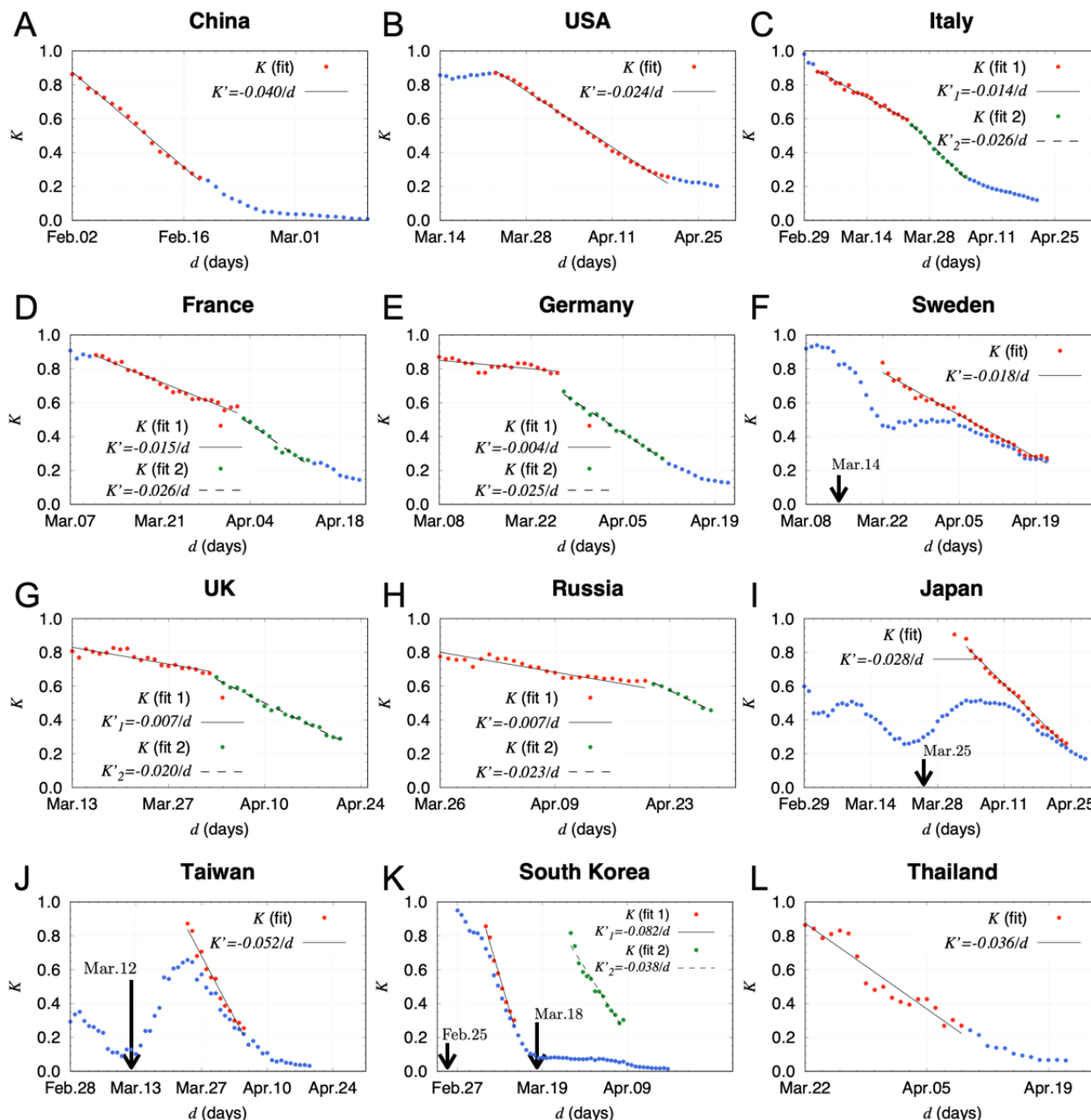
Discussion

Principal Findings and Interpretation of Results

As a new indicator of the COVID-19 spread rate, we have proposed the K value that takes a constant value if the number of infected people grows exponentially. The K values started to decrease even in the early stage of the COVID-19 spread, prior to the implementation of containment policies by the analyses of the real data for various countries. This suggests that the number of infected people does not follow the exponential growth. The analyses also revealed approximate

linear decrease in the time evolution of the K value. The slope of the linear decrease differed from country to country. The characteristic steep slope in East and Southeast Asian countries in the early stage implies the existence of immunological factors [11,12] that suppressed the spread of COVID-19.

Interpretation of the Linear K Transition From a Phenomenological Model

We have found that the K trajectory in the region of $K < 0.25$ was reproduced by a phenomenological model with the constant attenuation assumption. If $N(t)$ grows exponentially, the time

evolution of $N(t)$ is expressed as $N(t) = \exp(at)N(0)$, with a time-independent exponential constant a. However, the approximate linear decrease of the K value even in the early stage of the spread indicates that the constant a gradually decreased from the beginning. To introduce a small time dependence, we assumed that the constant a decreases exponentially, namely, $a(t) = \exp(-(1-k)t)\, a(0)$, or equivalently, $a(t) = \exp(-(1-k))\, a(t-1)$, with an attenuation factor k that is close to but less than 1. Under a condition of $(1-k) \ll 1$, the time evolution of $a(t)$ is approximated by $a(t) = ka(t-1) = k^t a(0)$, which leads to $N(t+1) = \exp(k^t a(0))N(t)$. The model calculations under this constant attenuation assumption showed that the K trajectory can be approximated by a first order linear function of t in a wide range $(0.25 < K < 0.9)$, with the slope K' being related to the attenuation factor k by the simple equation $k = 1 + 2.88K'$ (see Multimedia Appendix 1). $N(t)$ then follows the Gompertz curve [13,14], which is consistent with our

constant attenuation assumption. It was also confirmed that the simple equation $k = 1 + 2.88K'$ is derived analytically from the Taylor expansion at $K=0.5$ [13].

To test the long-term reliability and the validity of the assumption, we compared the model calculations with real data for Japan, France, Germany, and the United States. Using the k value calculated from K' obtained by a straight line fit in $K>0.25$, the K values below 0.25 were well reproduced by the model calculations for Japan, France, and Germany, as shown in Figure 2. This gives further indication that the total number of the infected people follows the Gompertz curve under fixed conditions for a long period. On the other hand, the data points started to deviate from the model calculation for the United States around April 18, 2020. The deviation was due to outbreaks in US states where COVID-19 spread was not serious in early April.

Figure 2. K values calculated by a phenomenological model with the constant attenuation assumption. (A) Model calculations and real data points for Japan. The slope K' was obtained by a linear fit in the range of $0.25 < K < 0.9$ using red points. The solid curve in the range of $K < 0.25$ is estimated with the constant attenuation factor k calculated from K'. The reference date was set for March 25, 2020. (B) Model calculations and real data points for France. (C) Model calculations and real data points for Germany. (D) Model calculations and real data points for the United States.
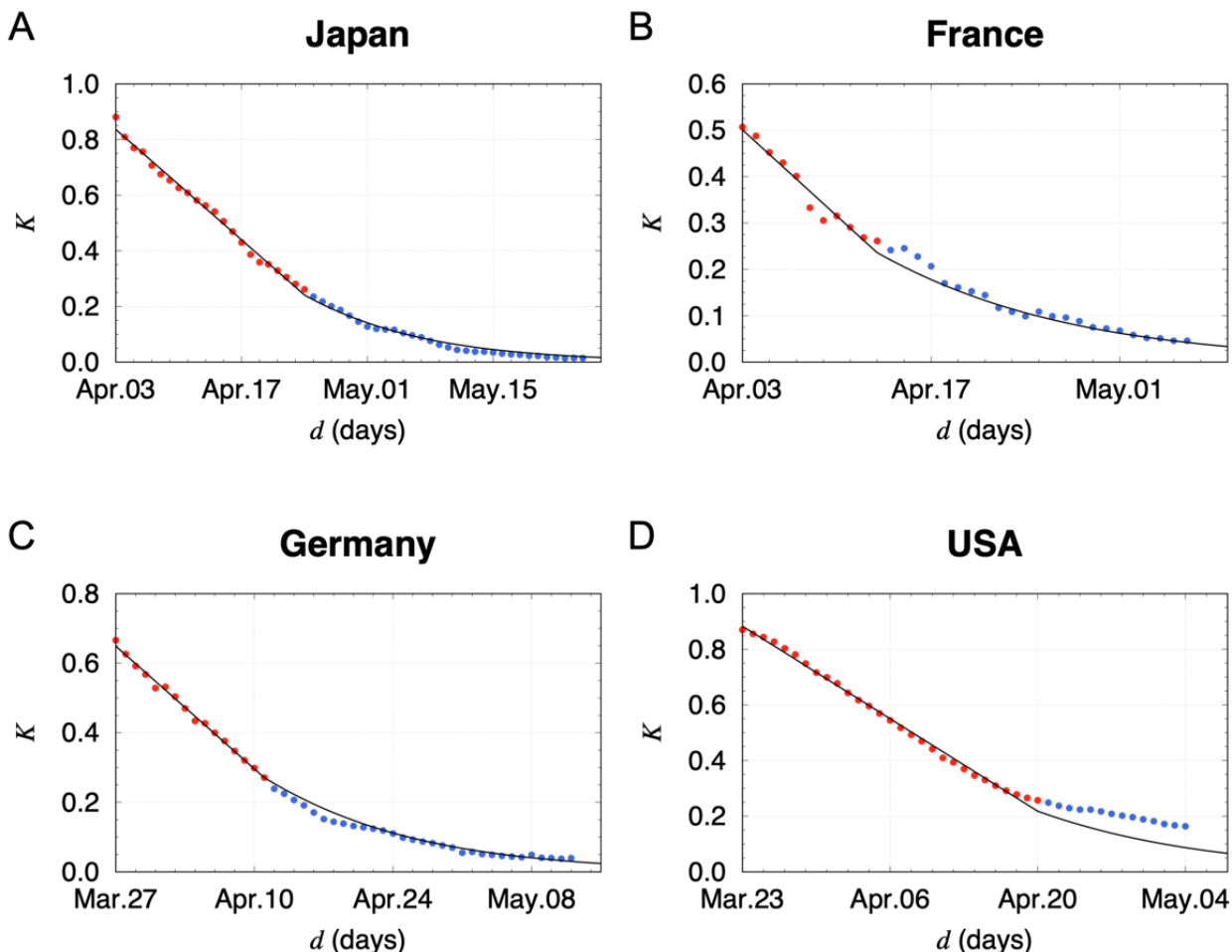
Interpretation of the Linear K Transition From a Susceptible-Infected Epidemic Model

To understand the behavior of the K and the slope K', we analyzed the public data of COVID-19 for the United States and Japan, employing the susceptible-infected (SI) model [10], which consists of the infected people (I) and susceptible people

($S = N - I$), with N being the final number of infected people. The SI model has the same time evolution of $I(t)$ as the SIR model in the early stages of the epidemic. Both models predict the exponential growth of $I(t)$.

In the United States and Japan, several changes in the slope of the K were found; it is natural to extend the standard SI model

to have several sources of COVID-19. The K is a monotonically decreasing function if a single source is taken into account like the standard SI model. The number of infected people with respect to each source i is described as:


The model parameters β and N_i control the spreading speed of COVID-19 and the final number of infected people caused by the source i, respectively. The analytic solution for I_i is found as $I_i(t) = (N_i/2)\,[1 + \tanh(\beta/2(t - t^c_i))]$, with t^c_i denoting the peak time of the infection spreading. The total number of infected people in a country at time t are then obtained as and finally reach As reported by Ranjan [15], each N_i should be much smaller than the countries' population sizes, so we chose N_i as fit parameters. The optimal solution of the parameter set $P = (\beta, \{N_i\}, \{t^c_i\})$ together with the number of sources n_{src} is found by minimizing the weighted mean squared deviation $L(P) = \sum_{d \in D}(I(d) - N(d))^2/N(d)$ in the fit range D. Here, $N(d)$ denotes the actual data of the total number of infected people on the day d. We chose D to be from February 23 to April 20, 2020 (58 days), for Japan and March 15 to April 28 (45 days) for the United States. We found the optimal number of sources for both countries as $n_{src}=4$, and the resulting parameters are given by

 = (.32, 142k, 303k, 279k, 365k, 03/26, 04/04, 04/14, 04/24)

for the United States and (.24, 0.2k, 0.7k, 3.1k, 8.7k, 02/22, 03/10, 04/03, 04/14) for Japan. Equipped with these parameters, the number of daily new cases dI_i/dt and the K value are presented in Figure 3.

For the United States, the first peak position (March 26, 2020) clearly corresponded to the date of the change in the slope K'. This shows that the K starts to decrease when the first peak out takes place. The other sources contributed to keep the K linear as superposition. For Japan, there were two peaks in the actual data of the K, and they started to decrease on March 12 in the first peak and on April 11 in the second one. The peak positions in the model agree with these dates. Moreover, the third peak position in the model coincides with the date when the K saturates at 0.5. Comparing with the model analysis of the United States, it is worth mentioning that the parameter β was smaller in Japan than that in the United States. Nevertheless, the steep gradient of K was observed in Japan. This observation is understood by the deference of the number of infection sources contributing to the K in both countries. In Japan, the first two and last two sources are well separated in time so that only the last two affect the decrease of the K as superposition. However, in the United States, all four sources contributed to it. The results show that K plays a crucial role to understand how the infection spreads. The linearity of the K value is not trivial but is most likely to be caused by several consecutive infectious explosions.

Figure 3. The model results of the total number of infected people and the K value in (A) the United States and (B) Japan. (A) The fit results are given in the top figure. The K value is represented together with the daily new cases dI_i/dt calculated in the susceptible-infected model at the bottom. The peak positions tci caused by the source i ($i=1,2,3,4$) are shown by the vertical allows. (B) Same as (A) for the case of Japan. The reference date of February 19 was kept for the K values.
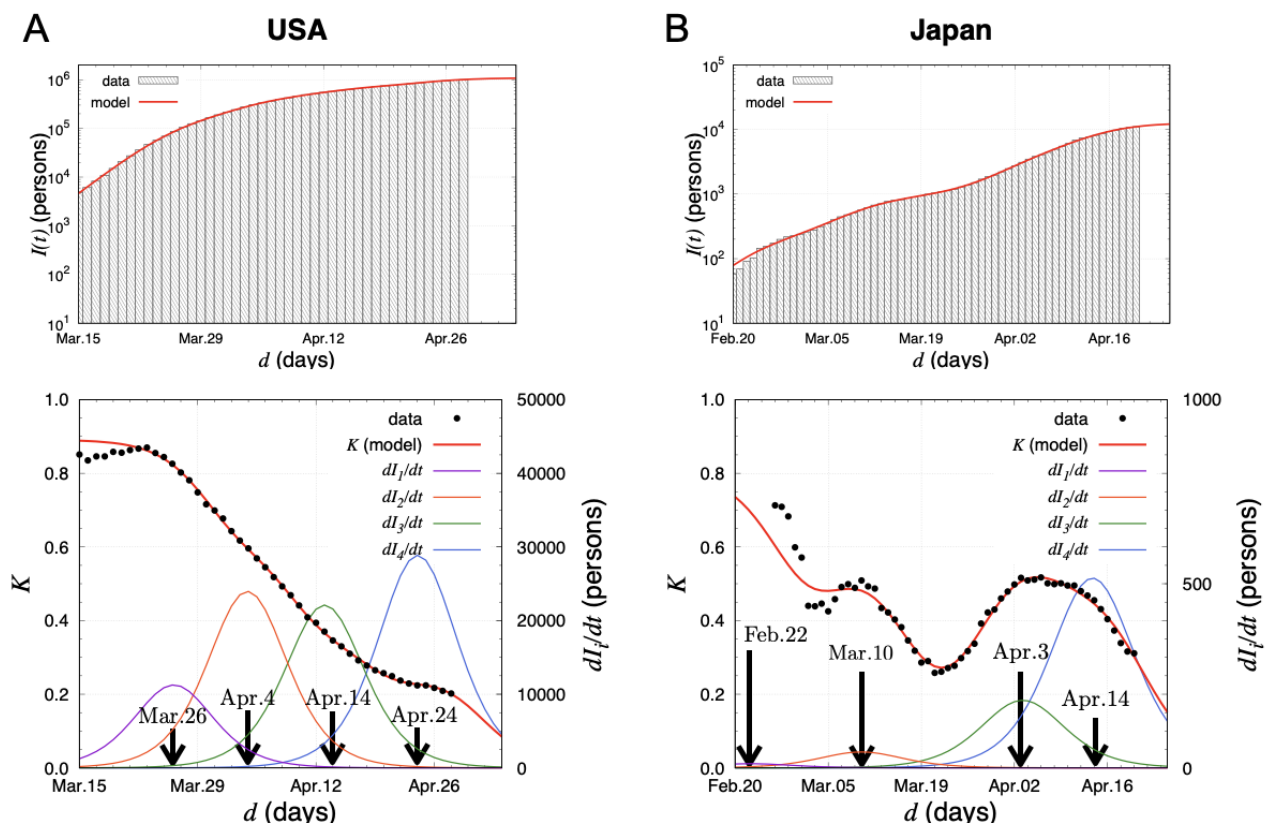
Conclusions

The analyses revealed that the time evolution of the K values in the region of $0.25 < K < 0.90$ can be universally approximated by a straight line in many cases, indicating that the total number of infected people follows the Gompertz curve. This finding will be helpful to improve and refine epidemiological models of COVID-19. The slope K' of the straight line was different from country to country, and the regional difference in the slope seemed to reflect both social and immunological circumstances for each country. Focusing on the change in the value of K will help to improve and refine epidemiological models of COVID-19 and other infectious diseases with the same tendency as COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Novel indicator of change in COVID-19 spread status.
[DOCX File , 103 KB - jmir_v22i11e20144_app1.docx]

References

1. Interactive web-based graphs of coronavirus disease 2019 cases and deaths per population by country. Our World in Data. 2020. URL: https://ourworldindata.org/coronavirus-source-data [accessed 2020-06-26]
2. US historical data. The COVID Tracking Project. 2020. URL: https://covidtracking.com/data/us-daily [accessed 2020-06-26]
3. COVID-19 special site. Nippon Television Network Corporation. URL: https://www.news24.jp/archives/corona_map/index2.html [accessed 2020-06-26]
4. Idogawa M, Tange S, Nakase H, Tokino T. Interactive web-based graphs of coronavirus disease 2019 cases and deaths per population by country. Clin Infect Dis 2020 Jul 28;71(15):902-903 [FREE Full text] [doi: 10.1093/cid/ciaa500] [Medline: 32339228]
5. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Royal Soc A 1927;115(772):700-721. [doi: 10.1098/rspa.1927.0118]
6. Ma J. Estimating epidemic exponential growth rate and basic reproduction number. Infect Dis Model 2020;5:129-141 [FREE Full text] [doi: 10.1016/j.idm.2019.12.009] [Medline: 31956741]
7. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. Proc Natl Acad Sci U S A 2020 Jul 21;117(29):16732-16738 [FREE Full text] [doi: 10.1073/pnas.2006520117] [Medline: 32616574]
8. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 2020 May 15;368(6492):742-746 [FREE Full text] [doi: 10.1126/science.abb4557] [Medline: 32269067]
9. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: a review. Phys Life Rev 2016 Sep;18:66-97 [FREE Full text] [doi: 10.1016/j.plrev.2016.07.005] [Medline: 27451336]
10. Hethcote HW. Qualitative analyses of communicable disease models. Math Biosciences 1976 Jan;28(3-4):335-356. [doi: 10.1016/0025-5564(76)90132-2]
11. Sala G, Chakraborti R, Ota A, Miyakawa T. Association of BCG vaccination policy with prevalence and mortality of COVID-19. MedRxiv. Preprint posted online May 17, 2020 . [doi: 10.1101/2020.03.30.20048165]
12. Iwasaki A, Grubaugh N. Why does Japan have so few cases of COVID-19? EMBO Mol Med 2020 May 08;12(5):e12481. [doi: 10.15252/emmm.202012481] [Medline: 32275804]
13. Akiyama Y. The K indicator epidemic model follows the Gompertz curve. Department of Computer Science School of Computing Tokyo Institute of Technology: Akiyama Lab. 2020. URL: http://www.bi.cs.titech.ac.jp/COVID-19/The_K_indicator_epidemic_model_follows_the_Gompertz_curve.html [accessed 2020-06-26]
14. Levitt M, Scaiewicz A, Zonta F. Predicting the trajectory of any COVID19 epidemic from the best straight line. medRxiv. Preprint posted online June 30, 2020 . [doi: 10.1101/2020.06.26.20140814] [Medline: 32607515]
15. Ranjan R. Estimating the final epidemic size for COVID-19 outbreak using improved epidemiological models. medRxiv. Preprint posted online April 20, 2020 . [doi: 10.1101/2020.04.12.20061002]

Abbreviations

SI: susceptible-infected

XSL•FO

RenderX

SIR: susceptible-infected-removed

XSL•FO
RenderX

<u>Original Paper</u>

Automated Smart Home Assessment to Support Pain Management: Multiple Methods Analysis

Roschelle L Fritz[1][*], BSN, MSN, RN, PhD; Marian Wilson[1][*], MPH, RN, PhD; Gordana Dermody[2][*], BSN, RN, PhD; Maureen Schmitter-Edgecombe[3][*], PhD; Diane J Cook[3][*], PhD

[1]College of Nursing, Washington State University, Vancouver, WA, United States

[2]School of Nursing and Midwifery, Edith Cowan University, Joondalup, Australia

[3]School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, United States

[*]all authors contributed equally

Corresponding Author:
Roschelle L Fritz, BSN, MSN, RN, PhD
College of Nursing
Washington State University
14204 NE Salmon Creek Avenue
Vancouver, WA, 98686-9600
United States
Phone: 1 3605469623
Email: shelly.fritz@wsu.edu

Abstract

Background: Poorly managed pain can lead to substance use disorders, depression, suicide, worsening health, and increased use of health services. Most pain assessments occur in clinical settings away from patients' natural environments. Advances in smart home technology may allow observation of pain in the home setting. Smart homes recognizing human behaviors may be useful for quantifying functional pain interference, thereby creating new ways of assessing pain and supporting people living with pain.

Objective: This study aimed to determine if a smart home can detect pain-related behaviors to perform automated assessment and support intervention for persons with chronic pain.

Methods: A multiple methods, secondary data analysis was conducted using historic ambient sensor data and weekly nursing assessment data from 11 independent older adults reporting pain across 1-2 years of smart home monitoring. A qualitative approach was used to interpret sensor-based data of 27 unique pain events to support clinician-guided training of a machine learning model. A periodogram was used to calculate circadian rhythm strength, and a random forest containing 100 trees was employed to train a machine learning model to recognize pain-related behaviors. The model extracted 550 behavioral markers for each sensor-based data segment. These were treated as both a binary classification problem (event, control) and a regression problem.

Results: We found 13 clinically relevant behaviors, revealing 6 pain-related behavioral qualitative themes. Quantitative results were classified using a clinician-guided random forest technique that yielded a classification accuracy of 0.70, sensitivity of 0.72, specificity of 0.69, area under the receiver operating characteristic curve of 0.756, and area under the precision-recall curve of 0.777 in comparison to using standard anomaly detection techniques without clinician guidance (0.16 accuracy achieved; $P<.001$). The regression formulation achieved moderate correlation, with $r=0.42$.

Conclusions: Findings of this secondary data analysis reveal that a pain-assessing smart home may recognize pain-related behaviors. Utilizing clinicians' real-world knowledge when developing pain-assessing machine learning models improves the model's performance. A larger study focusing on pain-related behaviors is warranted to improve and test model performance.

KEYWORDS

XSL•FO
RenderX

Introduction

More than 50 million US adults suffer from chronic pain, and 19.6 million experience high-impact chronic pain severe enough to interfere with daily living or work activities [1]. Guidelines from the Centers for Disease Control and Prevention favor nonopioid strategies to manage pain [2]; however, professionals treating pain have raised concerns that reducing access to opioids for the 18 million Americans using them for chronic pain will cause needless suffering [3]. Thus, there is a critical need to better understand how adults experiencing pain manage their symptoms at home. Artificial intelligence (AI) may afford the opportunity for observations leading to new understandings and improved home-based pain management. AI for health care has already afforded new perspectives [4] on automated assessments leading to novel and timely interventions [5]. Machine learning (ML) models are used in medical imaging [6,7], neurology [8], cardiology [9,10], pulmonology [11], nephrology [12,13], gastroenterology [14], pathology [15,16], health care informatics [17,18], and clinical decision support [5,19]. ML models capable of automated in-home assessments and alerts are also in the early stages of supporting individualized home-based interventions [20-24].

Social cognitive theory supports that mastering daily tasks is key to living with chronic pain [25]. Pain interference is best captured through observing physical and social changes in daily activities [26]. However, most pain assessments occur in clinical settings. Advances in smart home technology provide the opportunity for unobtrusive and continuous monitoring of daily activities [27]. Such monitoring offers sensor-based observation of activities, routines, and behaviors and could provide direct evidence of clinically relevant changes in daily routines [21,22], sleep [28], and socialization [22]. To date, smart home ML models have proven capable of differentiating behavior markers between groups [29,30], modeling characteristics of older adults' daily activities [20], recognizing dementia-related behaviors [31], predicting cognitive and mobility scores [20,32], forecasting behavior-based sleep and wake patterns [28], and recognizing health events such as falls [22], pneumonia [33], and depression [34].

A common limitation to developing ML health-behavior models is the time needed to provide real-world context (ground truth) for datasets. However, evidence exists that techniques accounting for small samples and low levels of consistent reporting can produce robust models [35,36]. While it may be difficult to predict pain experience trajectories, well-validated risk prediction models have identified individuals at risk for long-term pain [37]. Predictive models have been tested for low back pain [37], post-surgical cancer pain [38], and pain with dementia [39]. Statistical modeling has also been used to predict physical and psychological factors for long-term pain [40]; however, models have not yet been developed to identify pain-reducing behaviors. Modifiable behaviors (eg, exercise) remain poorly understood in the context of pain symptomology. We hypothesize that smart homes may assist with detecting behaviors that are influenced by pain (eg, sleep, socialization). However, data without contextual interpretation have little meaning in real-world situations and should be avoided in health care delivery [41,42]. Our smart home development methods illuminate one strategy for integrating clinical knowledge to support the development of a prototype pain-assessing smart home (PASH).

This multiple methods, secondary analysis used data from an ongoing longitudinal smart home study (2017-2021). The longitudinal study's ML models have already demonstrated the ability to recognize 30 activities of daily living (eg, sleeping, eating, entering or exiting home) with approximately 98% accuracy based on 3-fold cross validation [43]. The longitudinal study aims to train an ML model to recognize changes in health states in real-time in older adults with comorbidities. Participants are monitored for 1-2 years using smart home sensors (passive infrared motion, magnetic door use, light, temperature, and humidity) that are deployed in their current residence. No cameras or microphones are used. Secured date and time-stamp sensor data are collected, and the ML model labels daily activities in real-time [44]. Additionally, expert nurses with advanced practice skills conduct weekly health assessments of participants via telehealth or home visit. Individualized semistructured interviews are conducted to obtain information about potential changes in health status and behavior, and any health-related concerns are documented. Participants are asked to recall health changes (ie, health events) occurring in the 7 days prior to the nursing visit. Participants are asked: "How has your health been in the last week?" and "Did you have any particular days that weren't normal?" To elicit recall, questions are asked about each body system. For example, with a person having atrial fibrillation, the nurse might say: "I see your heart rate is X; have you had any issues with your heart since we last talked?" For sleep, they ask: "Were there any nights you didn't sleep well? Or went to bed late, or got up early? What was different about that day?" This line of questioning is designed to elicit information about the "how, what, when, where, and why" of health events. Vital signs, information about medication changes, new symptoms by body system (eg, neuro, cardiac), sleep quality, psychosocial status, functional status, and changes to daily routines (eg, all-day outings) are captured. A participatory approach is encouraged, and most participants keep a daily journal to help with recall, though this is not required. Health events are documented and coded by medical type (eg, neuro, cardiac, respiratory). Nurses match health event and sensor data and interpret event start and stop times as well as pre-event and post-event activities and behaviors (ie, providing real-world context, clinical ground truth). To train PASH, we used these existing pain-related clinical and sensor-based data. More information about the role of nurses in the longitudinal study is available in the literature [42,45,46].

Although pain was not the focus of the longitudinal study (the focus was on chronic conditions), multiple episodes of pain were captured by nurses. These were available in the archived health assessment records and used for this secondary analysis. The purpose of this secondary analysis was to determine if ambient sensor-based data could be used to train an ML model that recognizes pain-related behaviors. The longitudinal study and secondary analysis were approved by the Washington State University Institutional Review Board.

Methods

Data Inclusion Criteria

To train an ML model to recognize pain-related behaviors, we conducted a secondary data analysis of historic longitudinal sensor-based data and semistructured, holistic nursing assessment interviews containing descriptions of pain. Historic records afforded a convenient and purposive sample [47] of 17 independent, community-dwelling smart home participants aged ≥55 years with ≥2 chronic conditions, living alone, without pets. Of these, 11 participants met the inclusion criterion of having at least one recorded pain event during the longitudinal study. For each participant, approximately 60,000 sensor events per month were available, totaling 720,000 per year. All data from passive infrared and door use sensors throughout 1-2 years' monitoring were included for training the model. Based on prior work developing techniques to train ML models using small sample sizes [35,36], we determined this amount of data exceeds ML training requirements for models using longitudinal data and that this amount of data would likely allow the model to capture aspects of pain. The health records of *each* participant contained data from approximately 50-100 nursing assessments (1-2 years of weekly visits). We extracted pain-related information from the nursing record, defining a "pain event" as any report of pain associated with a report of related behavior changes. We also included 1 week of sensor data surrounding the pain event for conducting qualitative analysis of the sensor data as well as several weeks of baseline activity data — these were weeks where the nurse noted no health changes had occurred. "Normal" weeks were compared to weeks containing health events to help illuminate the event in the sensor data. We included data showing short-term visitors, which is exhibited in the data as multiple sensors turning ON nearly simultaneously (within <0.01 seconds of each other and <3 meters apart).

Data Exclusion Criteria

We excluded personal health data unrelated to pain. When conducting qualitative analysis of the sensor-based data, we excluded sensor data outside of the week surrounding the pain event except the weeks chosen to represent baseline normal routines. We also excluded data showing extended stay visitors (ie, stays across multiple days and nights). More information on the nursing team's analytic methods, including data exclusion processes, is available in the literature [45,46]. For training the ML model, we excluded data from light, temperature, and humidity sensors.

Qualitative Analysis

Adapted Qualitative Descriptive Methods

We applied the Fritz method [45] when analyzing pain-related, sensor-based data to support our expert-guided approach. The Fritz method includes the parallel processing of qualitative health event data and associated sensor data for contextualizing health changes in sensor data, enabling the development of clinically accurate ground truth [45]. It is an analytic approach that uses qualitative data and traditions to make sense of sensor data. Nurses use subjective semistructured interviews, objective nursing assessments, medical records, and clinical knowledge of the human response to illness to understand participants' health events and daily routines that are represented in the sensor data. For each identified health event, 1 week of sensor data around the time of the event are reviewed. Abnormal behavior patterns (not aligning with known daily routines) are identified and verified by comparing them with selected baseline routine datasets from the 6 months surrounding the health event. A clearly exhibited health event includes changes to normal routines (eg, wake time, time out of home, time in bathroom). For more information, see [45,46].

Analyzing Sensor Data

Two nurse analysts trained in qualitative descriptive methods [48] and the Fritz method [45] separately analyzed all 27 pain events. Each analyst used the nursing record, which included associated dates and times, diagnosis, and a summary of daily routines, to determine each pain event's timing and activities. The first round of analysis was conducted by Nurse Analyst A as part of the ongoing longitudinal study. The second round of analysis was conducted by Nurse Analyst B for the current substudy. After the second round of analysis was complete, Nurse Analysts A and B met to discuss potential themes. Figure 1 illustrates the qualitative analytic process that preceded computer processing of the data.

Figure 1. Multiple methods analytic processing of qualitative and quantitative data for a clinician-guided approach to machine learning (ML).
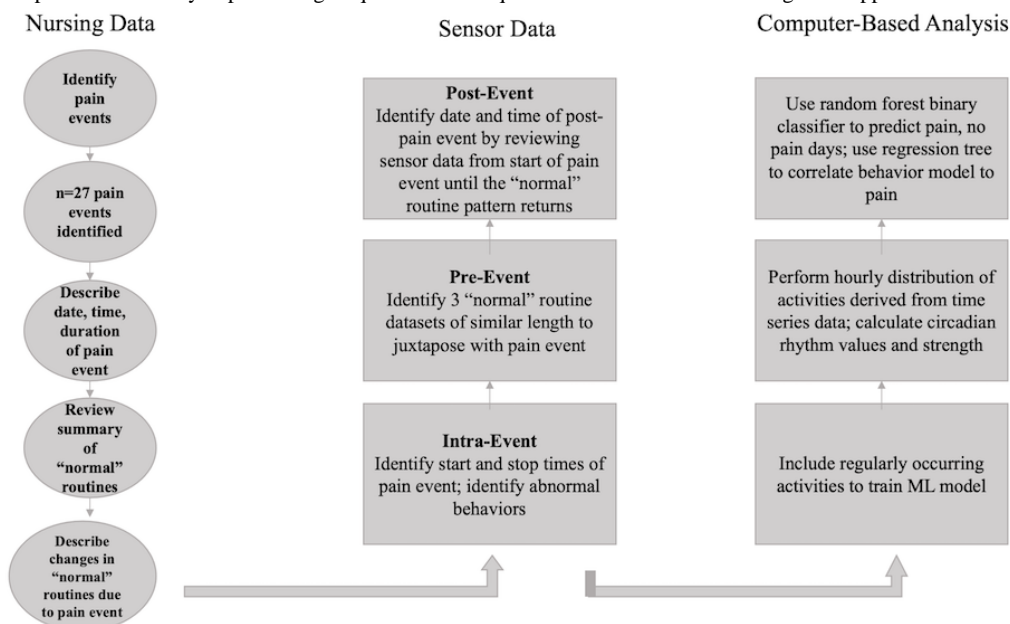

To analyze each pain event, the nursing record was reviewed first; then, the document summarizing the participant's daily routines was reviewed, and the associated sensor data were downloaded from a secure database. The day(s) of the event and a minimum of 1 week of sensor data surrounding the event (5 days before and 2 days after) were downloaded. If a pain event lasted 7 days, 2 weeks' data were analyzed: the week-long event plus 5 days before and 2 days after. Additionally, a minimum of 3 weeks representing "normal/baseline" sensor data were downloaded. We defined a normal week as any week where (1) the nursing record reported the participant had a normal week (ie, they said nothing had changed, or they said they felt good), (2) overnight visitors were not present, and, (3) it was not a holiday. Nurse analysts then looked at activity timing and duration, activity sequences, and the amount of data produced by each sensor on the day(s) of the event. We determined the timing and duration of activities by observing the sensor label (eg, bed, recliner) and the time of day a sensor transition occurred (eg, between the bedroom and living room). We determined activity sequences by observing the order in which various sensors' ON signals appeared. We determined the amount of data by calculating the total number of consecutive ON and OFF signals of a single sensor or cluster of sensors (eg, bed, general bedroom area). For example, insomnia or restlessness in the night was observed in the sensor data as an increase in the total number of consecutive bed sensor ON signals or the intermixing of other sensors' ON signals (eg, kitchen sensors) instead of sensor quietness (ie, absence of sensor ON signals, representing sleep). Once the pain event was clearly identified in the sensor data (eg, significant change in sleep behavior), baseline data of routine behaviors from the surrounding weeks and months were compared to the behavior anomalies associated with the pain event. A minimum of 3 datasets per event representing normal routines were captured for comparison; however, nurse analysts continually expanded their review of the sensor data surrounding the pain event until they were satisfied that reviewing more data would not produce additional understandings.

Developing Themes

Figure 2 illustrates the influence of abnormal behaviors (associated with experiencing pain) on emerging qualitative behavioral themes. Patterns in the sensor data representing pain-related behaviors became apparent when we compared sensor-based data across all 27 different pain events and across all participants (N=11). Emerging patterns were designated as behavioral themes. Some emerging themes incorporated related ideas; so, they were clustered together and assigned a larger, overarching idea that became the major theme. For example, the overarching idea of "Sleep" subsumed minor themes that regarded characteristics of sleep like the timing of sleep (eg, bedtime, wake time), length and quality of sleep (with and without interruption), and sleep location. Major themes were chosen based on 2 criteria: The theme was present across multiple events, and both nurses thought the theme was clinically relevant. No themes were dropped during the process of moving from minor to major themes. A final comprehensive re-review of transcripts did not reveal new themes. Both nurse analysts agreed on the chosen themes, and each nurse analyst's ideas were equally valued and addressed. Major themes represent activities that persons experiencing pain will likely alter. Such knowledge, emerging from sensor-based observations and clinical interpretation, supported the training of PASH.

Figure 2. Abnormal pain-related behaviors observed in the sensor data, with 6 overarching themes.
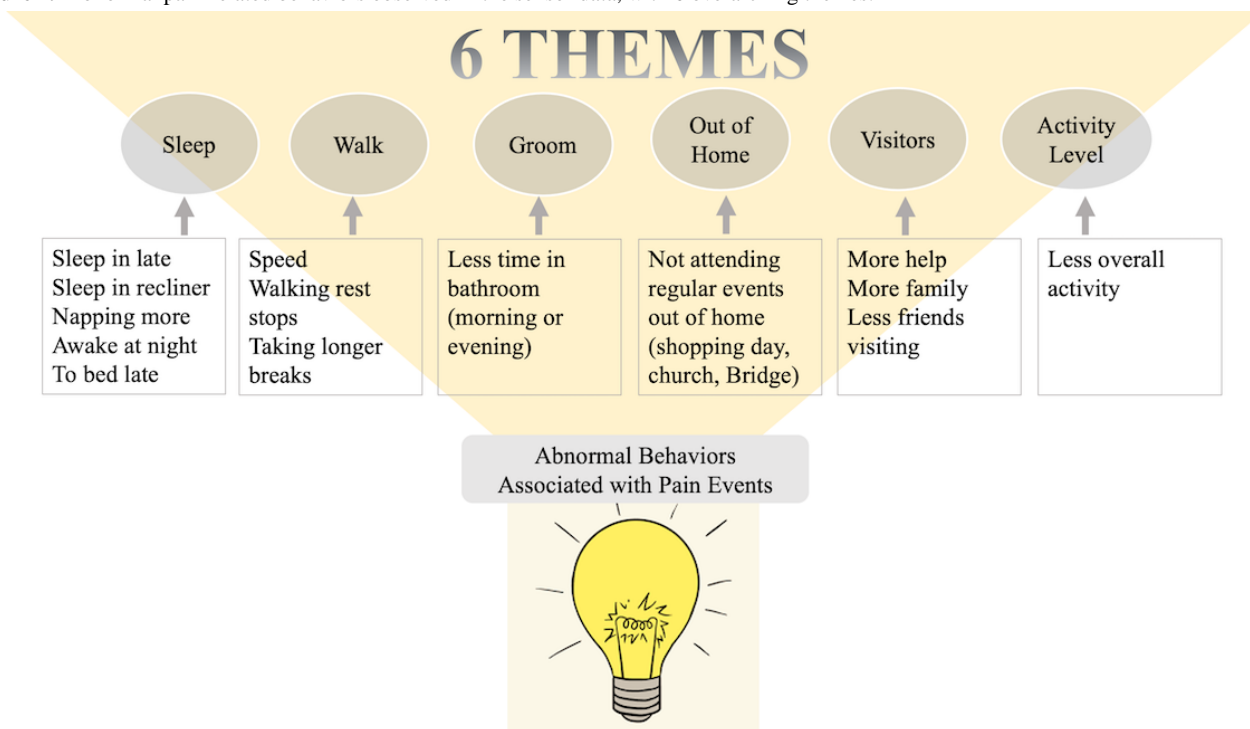
Quantitative Analysis

The same 11 smart homes were included for both qualitative and quantitative analyses. Only 2 sensor types (passive infrared motion, door use) were processed for the quantitative portion of the study. Pain events varied in length from 1 to 15 days (mean 7.85 days); for each pain event, an equivalent set of relatively pain-free data was included for each participant's pain event dataset. All collected smart home sensor events had already been automatically labeled with corresponding normal daily activities using our activity recognition algorithm that had previously been trained on 30 homes (not analyzed as part of this study). In each of these prior homes, a non-clinical annotator provided ground truth labels for 2 months of sensor data. From activity-labeled sensor data for the 11 homes included in this study, 550 behavior markers were extracted. A set of activities that occur with sufficient regularity in the homes (at least once a day on average) was selected to inform the ML models. The activity categories were Bathe, Bed-Toilet Transition, Cook, Eat, Enter Home, Leave Home, Personal Hygiene, Relax, Sleep, Take Medicine, Wash Dishes, Work, and Other Activity. The behavior markers corresponded to statistical measures of mean, median, standard deviation, maximum, minimum, zero crossings, mean crossings, interquartile range, skewness, kurtosis, and signal energy. Each measure was applied to time series data indicating hourly distributions for overall activity level (measured as the number of motion sensor events), hourly distributions over home locations, and hourly distributions over activity classes. Additionally, behavior markers were computed that indicate daily schedule regularity and 24-hour circadian rhythm values. The regularity value calculates the normalized difference (in location and in activity) between the same hours across all of the days in the sample. The circadian rhythm strength was calculated using a periodogram, which is an estimate of the spectral density of a signal (in this case, activity

level). The periodogram identifies the strength of the frequencies that explain variations in time series data. We quantified circadian rhythm as the normalized strength of the frequency that corresponds to 24 hours.

We designed several ML approaches for detecting pain events from sensor-derived behavior markers. First, we employed a random forest binary classifier with 150 trees to predict whether a day was part of a pain event (positive class) or pain-free (negative class). Second, we trained a regression tree on the data to determine correlation between a model of behavior markers and pain events. For comparison with our clinician-trained ML approaches, we employed an isolation forest (iForest) anomaly detection algorithm with 100 estimators to detect pain days. Unlike the random forest and regression trees, iForest did not use any clinician guidance in detecting anomalies that may indicate days for which the participant was experiencing pain. Finally, we trained a decision tree algorithm based on the positive and negative instances to determine which behavior markers provided the greatest distinction between pain days and those that were pain-free. We did not train the algorithm to differentiate between pain subclasses due to the small number of events in each group (acute, flare); however, our clinical team's early work determining these subgroups using sensor-based data positions us to explore subgroup comparisons in the future with larger sample sizes.

Results

Table 1 shows the sample characteristics of the participants and number and duration of acute and flair pain events. A total of 11 older adult participants, aged 68-92 years, were included in this secondary data analysis. All participants were community-dwelling and living independently while being

monitored with smart home sensors, except 1 participant who moved to assisted living during the study.

Table 1. Demographic characteristics of the study sample (N=11).

Characteristics	Results
Age (years), mean (range)	85.72 (69-92)
Biologic sex, n (%)	
Female	9 (82)
Male	2 (18)
Marital status, n (%)	
Married	1 (9)
Divorced	1 (9)
Widowed	9 (82)
Education, n (%)	
High school (some or graduated)	2 (18)
College (some or graduated with Bachelor's)	7 (64)
Graduate School (Master's or Doctorate)	1 (9)
Independence, n (%)	
Living independently	11 (100)[a]
Living alone	11 (100)
Uses assistive personnel (excluding housekeeping)[b]	2 (18)
Uses a housekeeper	10 (91)
Uses assistive equipment[c]	4 (36)
Pain event (n=27), n (%)	
Acute (duration=0.25-14 days; mean duration 6.8 days)	8 (30)
Flare (duration=2-8 days; mean duration 6.6 days)	19 (70)

[a]One participant entered assisted living after 17 months in the study.

[b]Assistive tasks: donning compression stockings (independent participant), medication administration, showering (assisted-living participant).

[c]Equipment: 4-wheeled walker, electric scooter, prosthetics.

Qualitative

We found 13 pain-related behaviors: no exit home, decreased time out of home, visitors, sleep location, time of sleep, length of sleep, night time sleep interruption, change in total sleep hours (increase or decrease), sleep quality (body movement during sleep — tossing and turning), grooming, walking speed, change in walking pattern (short bursts, long rests), overall activity in 24 hours. From the sensor activity patterns, 6 themes representing all 13 behaviors emerged (Figure 2): Sleep, Walking, Grooming, Time Spent Out of Home, Visitors, Overall Activity Level. Themes represented pain-related behaviors. Table 2 shows major and minor theme characteristics, the types of pain reported by participants (noted by nurses), and the type of sensor, location, and sensor combinations informing qualitative interpretations. Themes generally aligned with the 13 activity categories that the ML algorithm recognized, and each of the 6 themes was represented in the subset of 30 activities (from the prior homes) included for training PASH. For example, the qualitative theme of "Time Spent Out of Home" was represented in PASH's ML model as "Enter Home" and "Leave Home."

Table 2. Themes representing pain-related behaviors.

Themes (activity attributes)	Participant-reported pain	Nursing report of participants' behaviors	Sensors informing themes	Meaningful sensor combinations	Sensor type
Sleep (rhythm, length, location, quality)	Neck, leg, knee, hip pain	"slept in," "moved to recliner," "increased nap time," "awake in night"	General bedroom, bed, recliner	<Bed-Toilet>, <Recliner-Toilet>, <Bed-Recliner>, <Bed-Bedroom>, <Bed-Kitchen>	Passive infrared (PIR)
Walking (speed, rests, breaks)	Fall, leg, knee, hip pain; chest pain	"shortness of breath," "resting more frequently when walking"	Bedroom, bed, recliner, toilet, hallway	<Recliner-Toilet>, <Bed-Toilet>, <Hallway-Hallway>	PIR
Grooming (done, not done)	Fall, abdominal pain	"has not showered for 2 days,"[a] "help with grooming"	Bathroom, bathroom sink	<General Bathroom Area-Bathroom Sink> (Quantity and duration)	PIR
Time spent out of home	Fall, neck, leg, knee, abdominal pain	"didn't go to Bridge night," "didn't go shopping"	Main entry	<Main Entry-Hallway>, <Absence of sensor events>	Magnetic contact (door use)
Visitors (Social, health workers)	Fall, neck, leg, abdominal pain	"now has home health," "caregiver at bedtime," "daughter visit for 3 days[a] to help"	General living room, recliner, kitchen sink, hallway, bathroom sink	Any 2 sensors with virtually concurrent ON signals (<0.01 seconds apart) located greater than 10 feet apart	PIR
Overall activity level (in 24 hours)	Fall; leg, hip, abdominal pain; chest pain	"didn't attend exercise class," "didn't feel like doing much, just laid on sofa," "mostly in bed for 2 days"[a]	All sensors	Total number of sensor events in 24 hours, room activity length and variety; account for time out of home	PIR, magnetic contact (door use)

[a]Nursing record contained actual dates of participant-reported pain events.

Sleep was the leading pain-related behavioral theme, accounting for 6 of the 13 described behaviors. Of the 11 participants, 8 (having 22 of 27 pain events [81%]) reported sleep changes that were observed in the sensor data. Leading observable sleep behaviors included location (more time in a recliner), timing (bedtime, wake time, napping), hours of sleep at night, and sleep quality. Six of 8 acute pain events (short-term pain not associated with underlying pain) affected sleep, resulting in 2 people spending more time in their recliner chair, 1 person experiencing decreased overall sleep (day and night) and an inability to sleep in bed, and 3 people with restless sleep (decreased sleep quality). Twelve of 19 flare pain events (exacerbation of underlying pain) affected sleep, observed as more time in bed during the day, earlier bedtimes, more time spent in recliners, and intermittent changes in sleep location across several months. Two walking characteristics were observed: Walking speed slowed, and the number or length of rest breaks increased. We observed this in the monthly Timed-Up-and-Go tests and by calculating the difference in the time it took to move between specific sensors (bed and toilet or recliner and toilet).

Grooming activities were observed by reviewing bathroom sink, bedroom, and bedroom closet sensor groupings. A lack of grooming recorded in the nursing record as "stayed in bed 3 days on Dilaudid" was seen in the sensor data as a decrease in overall time spent in the bathroom in the mornings and evenings. A lack of grooming appeared to be associated with pain intensity. For example, the nursing record reported that 1 participant said, "I just haven't felt like leaving the house or even showering. I mostly lay in bed all day." Other pain-related behavior modifications regarded participants' overall activity level (in 24 hours) and time spent out of the home. All acute pain events resulted in decreased overall activity: 2 reduced

their out-of-home activities, and 2 had more visitors. Pain-event interpretations, event start and stop times, ground truth annotations, and themes were communicated to the computer science team to support the training of PASH.

Quantitative

The random forest classifier that was used to distinguish pain from pain-free days yielded a 3-fold cross-validation classification accuracy of 0.70, sensitivity of 0.72, specificity of 0.69, area under the receiver operating characteristic curve value of 0.756, and area under the precision-recall curve of 0.777. We use the term pain-free to distinguish between pain events and routine days; participants may not have actually been pain-free. To check the impact of clinicians' ground truth annotations and the expert-guided approach, we used standard anomaly detection techniques to determine classification accuracy without expert guidance. We used iForest to determine periods of time that were generally considered anomalous without clinician guidance. Using the detected anomalies (no clinician guidance) as indicators of pain events yielded a predictive accuracy of 0.16, a difference from the random forest (with clinician guidance) that was statistically significant ($P<.001$). Using the regression tree, a moderate correlation ($r=0.415$) was found between the behavior models and an indication of pain on the corresponding days.

The decision tree classifier created a greedy ordering of behavior features that best distinguish pain from pain-free days using information gain as the ordering criterion. These results highlighted 3 features that provided a high level of differentiation between the 2 classes: normalized overall activity level (lower for pain days), time spent in bed-toilet transition activities (higher for pain days), and time spent in a favorite chair (higher for pain days). Behaviors such as overall activity

level, walking speed (or time spent in bed-toilet transition), and time spent in a favorite chair (including sleeping there) were supported by both qualitative and quantitative analytic findings and may prove important to understanding pain experiences.

Discussion

Principal Findings

Our findings show preliminarily that ambient sensor-based data can be used to train an ML model to recognize pain-associated behaviors. These findings align with previous studies that indicate ML algorithms are capable of detecting behaviors that indicate a clinically relevant change in health status [21,23,32,33,49,50]. Unlike these previous studies, which primarily focused on associating behaviors with cognitive and functional health, we focused on behaviors exhibited by persons experiencing pain. While there is significant overlap in pain-related behaviors and other known health-related behaviors that machines can recognize, we uniquely discovered that one behavior not typically included in pain interference scales (yet recognized as physical activity sequences by ML algorithms) point to the existence of increased pain: decreased grooming. Importantly, almost all the qualitative themes (emerging from sensor-based data) align with behaviors that are already well-established with validated pain interference scales [26]. The exception is "Grooming." Grooming is not specifically identified as a factor in some of the most commonly used pain interference measurement scales (eg, the Brief Pain Inventory); yet, it could become an important target to assess pain populations when using sensor monitoring. This is an example of how smart home monitoring can generate new evidence-based information to support pain management.

Given that the standard anomaly detection techniques used to determine classification accuracy without expert guidance yielded a 16% classification accuracy compared with the expert-guided approach (predictive accuracy 70%), our findings support the ideas that (1) clinicians, such as nurses with frontline pain management experience, add value to the efficacy of the ML model and (2) PASH offers possibilities as a clinical tool for identifying pain-related behaviors. Though the model demonstrated a pragmatically low predictive accuracy (70%) for clinical applications and we did not ask it to discriminate between pain subclasses (acute, flare), it performed quite well given the small participant sample size (N=11) and small number of captured pain events (N=27). PASH's accuracy could improve given a larger participant sample size and greater number of pain-related training events interpreted by clinicians.

The question of whether all pain leads to the same pain-related behaviors could be raised. Our smart home approach to pain management cannot yet determine the type of pain. Further, this approach to pain assessment cannot determine the source, severity, or location of a person's pain (eg, abdominal versus head). However, given that ML models are capable of recognizing clinically relevant behavior changes [21,29], it is reasonable to consider that ML models could alert when anomalous pain-related behaviors occur regardless of which behavior the model chooses to prioritize or the pain characteristics. Based on our preliminary findings, it is also

plausible that, with larger samples, ML models could be trained to alert on unrealized pain-related behaviors. Such alerts, based on naturalistic real-time data of persons experiencing and attempting to manage pain, would be of significant value to clinicians seeking to perform early interventions using minimal pharmacologics. ML affords this opportunity while also individualizing pain context and offering the potential to discover new perspectives on pain. Randomized controlled trials supporting current interventions do not account for individual differences in pain experiences. Rather, they focus on average pain responses, leaving persons who are outliers without optimal care. Learning how individuals with pain uniquely express their pain moment-by-moment could lead to novel understandings of pain and afford the opportunity to provide effective, precise interventions. To achieve such precision, PASH would benefit from adding ecological momentary assessment, an in-the-moment data capture technique for naturalistic settings [51] now regarded as the most accurate method for capturing real-world pain [52,53].

Clinical Implications

PASH offers clinicians a more objective, data-driven way of knowing about pain. PASH could be of benefit to nonverbal or cognitively impaired individuals [54]. Developing consistent, reliable, objective pain measurements; detecting patterns in behaviors and activities that exacerbate or relieve pain; and accurately capturing responses to medications and other pain treatments are potential scientific discoveries that could be realized using an AI system like PASH. In addition, PASH could add new sensor-based evidence of biopsychosocial pain components and facilitate the combining of traditional and new data to augment and support clinical assessments (with reduced bias) and clinical decision making. We recognize that it is impossible to know when an acute pain condition may transition into chronic pain; yet, part of our enthusiasm about this work is the future possibility of detecting minor and important changes in this regard using sensor data. This would be an important contribution to pain science.

Privacy and data security have been identified as primary concerns for older adults considering the use of a smart home. As ubiquity of data collection expands to the home environment and is integrated into the delivery of health care, considerations for data security are needed so risks for data breaches are mitigated and identities of vulnerable persons, such as persons in pain, cannot be stolen or easily reconstructed.

Bias

Clinicians and software developers can potentially introduce bias into ML models [5,14,55]. Assumptions and generalized perspectives regarding subpopulations and disease progression tendencies incorporated in a clinician's belief structure (overtly or inadvertently) over time could impact the reporting of ground truth. Likewise, software developers' assumptions (recognized or unrecognized) have potential to impact design approaches and perceived end-user wants and needs. To avoid algorithm bias, clinicians and computer scientists as well as study participants need to be a diverse group of humans [56]. To accomplish this, intentionality toward diversity in all aspects of the design loop is required.

Limitations

This project was limited by the larger study's design and the purposive sample size (number of participants, number and type of pain events; both convenience samples). Using historic datasets not specifically collected for discovering pain-related behaviors as well as the small numbers of participants and pain events limit generalizability and make the model susceptible to overfit. Reliance on participants' weekly recall of health changes, including pain, limits accuracy of sensor data interpretations. Not all qualitative and quantitative components of this multiple methods, secondary analysis aligned. We do not make an attempt to differentiate between depression and pain-related behaviors in this study and acknowledge the synergy between these conditions. Significant time and effort are required for the expert-guided approach to ML, which potentially limits scalability. Larger and more diverse samples, a longitudinal design, and use of an ecological momentary assessment for data collection are needed. Using ML to identify and discriminate between pain phenotypes would be of benefit to providers and patients. PASH should be tested in a prospective study to identify true and false positive ratios. Additionally, PASH needs to be trained to accommodate a multiperson household.

Conclusion

Innovative monitoring and treatment options are needed to support persons experiencing chronic pain, their caregivers, and the health care professionals working alongside them to improve their quality of life and health outcomes. Our findings suggest that smart homes using AI monitoring tools are well-positioned to become pragmatically useful at detecting clinically relevant pain-related behaviors or relevant changes in those behaviors. Using smart homes to provide automated pain assessment and intervention could alleviate some of the pressure on patients and clinicians working in the pain management health care arena. Concrete, objective data demonstrating how people with pain are affected and how they self-manage painful conditions at home could be used to deepen understandings and innovate solutions. Leveraging such technologies for health care delivery should be done intentionally, and clinicians should participate in technology development studies to interpret data, provide meaningful context, and illuminate meaningful use possibilities in all phases of development.

Conflicts of Interest

None declared.

References

1. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]

2. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm. Rep 2016 Mar 18;65(1):1-49. [doi: 10.15585/mmwr.rr6501e1]

3. Darnall B, Juurlink D, Kerns R, Mackey S, Van Dorsten B, Humphreys K, et al. International Stakeholder Community of Pain Experts and Leaders Call for an Urgent Action on Forced Opioid Tapering. Pain Med 2019 Mar 01;20(3):429-433 [FREE Full text] [doi: 10.1093/pm/pny228] [Medline: 30496540]

4. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial Intelligence in Clinical Health Care Applications: Viewpoint. Interact J Med Res 2019 Apr 05;8(2):e12100 [FREE Full text] [doi: 10.2196/12100] [Medline: 30950806]

5. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books, Inc; 2019.

6. Odaibo D, Zhang Z, Skidmore F. Detection of visual signals for pneumonia in chest radiographs using weak supervision. SoutheastCon IEEE 2019:1-5. [doi: 10.1109/southeastcon42311.2019.9020631]

7. Kalpathy-Cramer J, Freymann J, Kirby J, Kinahan PE, Prior FW. Quantitative Imaging Network: Data Sharing and Competitive AlgorithmValidation Leveraging The Cancer Imaging Archive. Transl Oncol 2014 Feb;7(1):147-152 [FREE Full text] [doi: 10.1593/tlo.13862] [Medline: 24772218]

8. Regalia G, Onorati F, Lai M, Caborni C, Picard RW. Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands. Epilepsy Res 2019 Jul;153:79-82. [doi: 10.1016/j.eplepsyres.2019.02.007] [Medline: 30846346]

9. Rajpurkar P, Hannun AY, Haghpanahi M. Cardiologist-level arrhythmia detection with convolutional neural networks. 2017. URL: http://arxiv.org/abs/1707.01836 [accessed 2020-11-04]

XSL•FO

RenderX

10. Clifford G, Liu C, Moody B, Lehman LH, Silva I, Li Q, et al. AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. 2017 Presented at: 2017 Computing in Cardiology (CinC); Sept 24-27, 2017; Rennes, France p. 1-4. [doi: 10.22489/cinc.2017.065-469]

11. Topalovic M, Das N, Burgel P, Daenen M, Derom E, Haenebalcke C, Pulmonary Function Study Investigators. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. Eur Respir J 2019 Apr 14;53(4):1801660. [doi: 10.1183/13993003.01660-2018] [Medline: 30765505]

12. Niel O, Boussard C, Bastard P. Artificial Intelligence Can Predict GFR Decline During the Course of ADPKD. Am J Kidney Dis 2018 Jun;71(6):911-912. [doi: 10.1053/j.ajkd.2018.01.051] [Medline: 29609979]

13. Lemley KV. Machine Learning Comes to Nephrology. JASN 2019 Sep 05;30(10):1780-1781. [doi: 10.1681/asn.2019070664]

14. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]

15. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. IEEE Trans. Med. Imaging 2017 Jul;36(7):1550-1560. [doi: 10.1109/tmi.2017.2677499]

16. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019 Aug 15;25(8):1301-1309 [FREE Full text] [doi: 10.1038/s41591-019-0508-1] [Medline: 31308507]

17. Roberts A, Gaizauskas R, Hepple M, Davis N, Demetriou G, Guo Y, et al. The CLEF corpus: semantic annotation of clinical text. AMIA Annu Symp Proc 2007 Oct 11:625-629 [FREE Full text] [Medline: 18693911]

18. Rehm G, Kuhn B, Lieng M, Cortes-Puch I, Nguyen J, Guo EC, et al. An analytic platform for the rapid and reproducible annotation of ventilator waveform data. bioRxiv. 2019. URL: https://www.biorxiv.org/content/10.1101/568386v1 [accessed 2020-11-04]

19. Albright D, Lanfranchi A, Fredriksen A, Styler WF, Warner C, Hwang JD, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative. J Am Med Inform Assoc 2013 Sep 01;20(5):922-930 [FREE Full text] [doi: 10.1136/amiajnl-2012-001317] [Medline: 23355458]

20. Dawadi PN, Cook DJ, Schmitter-Edgecombe M. Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks. IEEE Trans. Syst. Man Cybern, Syst 2013 Nov;43(6):1302-1313. [doi: 10.1109/tsmc.2013.2252338]

21. Sprint G, Cook DJ, Fritz R, Schmitter-Edgecombe M. Using Smart Homes to Detect and Analyze Health Events. Computer 2016 Nov;49(11):29-37. [doi: 10.1109/mc.2016.338]

22. Sprint G, Cook D, Fritz R, Schmitter-Edgecombe M. Detecting health and behavior change by analyzing smart home sensor data. 2016 Presented at: IEEE International Conference on Smart Computing, SMARTCOMP; May 18-20, 2016; St. Louis, MO. [doi: 10.1109/smartcomp.2016.7501687]

23. Rantz M, Skubic M, Abbott C, Galambos C, Popescu M, Keller J, et al. Automated In-Home Fall Risk Assessment and Detection Sensor System for Elders. Gerontologist 2015 Jun 08;55 Suppl 1(Suppl 1):S78-S87 [FREE Full text] [doi: 10.1093/geront/gnv044] [Medline: 26055784]

24. Skubic M, Guevara RD, Rantz M. Automated Health Alerts Using In-Home Sensor Data for Embedded Health Assessment. IEEE J. Transl. Eng. Health Med 2015;3:1-11. [doi: 10.1109/jtehm.2015.2421499]

25. Keogh A, Tully MA, Matthews J, Hurley DA. A review of behaviour change theories and techniques used in group based self-management programmes for chronic low back pain and arthritis. Man Ther 2015 Dec;20(6):727-735. [doi: 10.1016/j.math.2015.03.014] [Medline: 25865062]

26. Wilson M. Integrating the concept of pain interference into pain management. Pain Manag Nurs 2014 Jun;15(2):499-505. [doi: 10.1016/j.pmn.2011.06.004] [Medline: 24882027]

27. LeBaron V, Hayes J, Gordon K, Alam R, Homdee N, Martinez Y, et al. Leveraging Smart Health Technology to Empower Patients and Family Caregivers in Managing Cancer Pain: Protocol for a Feasibility Study. JMIR Res Protoc 2019 Dec 09;8(12):e16178 [FREE Full text] [doi: 10.2196/16178] [Medline: 31815679]

28. Williams JA, Cook DJ. Forecasting behavior in smart homes based on sleep and wake patterns. THC 2017 Feb 21;25(1):89-110. [doi: 10.3233/thc-161255]

29. Cook DJ, Schmitter-Edgecombe M, Dawadi P. Analyzing Activity Behavior and Movement in a Naturalistic Environment Using Smart Home Techniques. IEEE J. Biomed. Health Inform 2015 Nov;19(6):1882-1892. [doi: 10.1109/jbhi.2015.2461659]

30. Sprint GL, Cook DJ, Fritz R. Behavioral Differences Between Subject Groups Identified Using Smart Homes and Change Point Detection. IEEE J. Biomed. Health Inform 2020:1. [doi: 10.1109/jbhi.2020.2999607]

31. Debes C, Merentitis A, Sukhanov S, Niessen M, Frangiadakis N, Bauer A. Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior. IEEE Signal Process. Mag 2016 Mar;33(2):81-94. [doi: 10.1109/MSP.2015.2503881]

32. Dawadi P, Cook D, Schmitter-Edgecombe M. Automated Cognitive Health Assessment From Smart Home-Based Behavior Data. IEEE J Biomed Health Inform 2016 Jul;20(4):1188-1194 [FREE Full text] [doi: 10.1109/JBHI.2015.2445754] [Medline: 26292348]

33. Rantz MJ, Skubic M, Popescu M, Galambos C, Koopman RJ, Alexander GL, et al. A New Paradigm of Technology-Enabled 'Vital Signs' for Early Detection of Health Change for Older Adults. Gerontology 2015 Nov 26;61(3):281-290 [FREE Full text] [doi: 10.1159/000366518] [Medline: 25428525]

34. Galambos C, Skubic M, Wang S, Rantz M. Management of Dementia and Depression Utilizing In- Home Passive Sensor Data. Gerontechnology 2013 Jan 01;11(3):457-468 [FREE Full text] [doi: 10.4017/gt.2013.11.3.004.00] [Medline: 24049513]

35. Hu N, Englebienne G, Lou Z, Krose B. Learning to Recognize Human Activities Using Soft Labels. IEEE Trans. Pattern Anal. Mach. Intell 2017 Oct 1;39(10):1973-1984. [doi: 10.1109/tpami.2016.2621761]

36. Dahmen J, Cook D. SynSys: A Synthetic Data Generation System for Healthcare Applications. Sensors (Basel) 2019 Mar 08;19(5):1181 [FREE Full text] [doi: 10.3390/s19051181] [Medline: 30857130]

37. Mukasa D, Sung J. A prediction model of low back pain risk: a population based cohort study in Korea. Korean J Pain 2020 Apr 01;33(2):153-165. [doi: 10.3344/kjp.2020.33.2.153] [Medline: 32235016]

38. Lötsch J, Sipilä R, Tasmuth T, Kringel D, Estlander A, Meretoja T, et al. Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy. Breast Cancer Res Treat 2018 Sep 6;171(2):399-411 [FREE Full text] [doi: 10.1007/s10549-018-4841-8] [Medline: 29876695]

39. Atee M, Hoti K, Hughes J. A Technical Note on the PainChek™ System: A Web Portal and Mobile Medical Device for Assessing Pain in People With Dementia. Front Aging Neurosci 2018;10:117 [FREE Full text] [doi: 10.3389/fnagi.2018.00117] [Medline: 29946251]

40. Meretoja TJ, Andersen KG, Bruce J, Haasio L, Sipilä R, Scott NW, et al. Clinical Prediction Model and Tool for Assessing Risk of Persistent Pain After Breast Cancer Surgery. JCO 2017 May 20;35(15):1660-1667. [doi: 10.1200/jco.2016.70.3413]

41. Forsythe DE, Hess DJ. Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence. Stanford, CA: Stanford University Press; 2001.

42. Dermody G, Fritz R. A conceptual framework for clinicians working with artificial intelligence and health-assistive Smart Homes. Nurs Inq 2019 Jan 12;26(1):e12267 [FREE Full text] [doi: 10.1111/nin.12267] [Medline: 30417510]

43. Aminikhanghahi S, Cook DJ. Enhancing activity recognition using CPD-based activity segmentation. Pervasive and Mobile Computing 2019 Feb;53:75-89. [doi: 10.1016/j.pmcj.2019.01.004]

44. Cook DJ, Crandall AS, Thomas BL, Krishnan NC. CASAS: A Smart Home in a Box. Computer 2013 Jul;46(7):62-69. [doi: 10.1109/mc.2012.328]

45. Fritz RL, Dermody G. A nurse-driven method for developing artificial intelligence in "smart" homes for aging-in-place. Nurs Outlook 2019 Mar;67(2):140-153 [FREE Full text] [doi: 10.1016/j.outlook.2018.11.004] [Medline: 30551883]

46. Fritz RL, Cook D. Identifying varying health states in smart home sensor data: An expert-guided approach. 2017 Presented at: World Multi-Conference of Systemics, Cybernetics Informatics: WMSCI; July 8-11, 2017; Orlando, FL.

47. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]

48. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]

49. Cook DJ, Schmitter-Edgecombe M, Jonsson L, Morant AV. Technology-Enabled Assessment of Functional Health. IEEE Rev Biomed Eng 2019;12:319-332. [doi: 10.1109/RBME.2018.2851500] [Medline: 29994684]

50. Robben S, Englebienne G, Krose B. Delta Features From Ambient Sensor Data are Good Predictors of Change in Functional Health. IEEE J. Biomed. Health Inform 2017 Jul;21(4):986-993. [doi: 10.1109/jbhi.2016.2593980]

51. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

52. Heapy A, Dziura J, Buta E, Goulet J, Kulas JF, Kerns RD. Using multiple daily pain ratings to improve reliability and assay sensitivity: how many is enough? J Pain 2014 Dec;15(12):1360-1365. [doi: 10.1016/j.jpain.2014.09.012] [Medline: 25283469]

53. Carlozzi NE, Schilling S, Freedman J, Kalpakjian CZ, Kratz AL. The reliability of end of day and ecological momentary assessments of pain and pain interference in individuals with spinal cord injury. Qual Life Res 2018 Nov 2;27(11):3003-3012 [FREE Full text] [doi: 10.1007/s11136-018-1952-y] [Medline: 30073468]

54. Schiavenato M, Craig K. Pain assessment as a social transaction: beyond the "gold standard". Clin J Pain 2010 Oct;26(8):667-676. [doi: 10.1097/AJP.0b013e3181e72507] [Medline: 20664341]

55. Yapo A, Weiss J. Ethical implications of bias in machine learning. 2018 Presented at: 51st Hawaii International Conference on System Sciences; January 3-6, 2018; Waikoloa Village, Hawaii p. 5365-5372 URL: http://hdl.handle.net/10125/50557 [doi: 10.24251/hicss.2018.668]

56. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]

Abbreviations

AI: artificial intelligence
iForest: isolation forest

ML: machine learning
PASH: pain-assessing smart home

XSL·FO
RenderX

Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach

Aaqib Shehzad[1*], BDS, MHI; Kenneth Rockwood[1,2,3*], MD; Justin Stanley[1*], BEng; Taylor Dunn[1*], MSc; Susan E Howlett[1,3,4*], PhD

[1]DGI Clinical Inc, Halifax, NS, Canada

[2]Geriatric Medicine Research Unit, Nova Scotia Health Authority, Halifax, NS, Canada

[3]Division of Geriatric Medicine, Dalhousie University, Halifax, NS, Canada

[4]Department of Pharmacology, Dalhousie University, Halifax, NS, Canada

[*]all authors contributed equally

Corresponding Author:
Kenneth Rockwood, MD
Geriatric Medicine Research Unit
Nova Scotia Health Authority
5955 Veterans' Memorial Lane
Halifax, NS, B3H 2E9
Canada
Phone: 1 9024738687
Fax: 1 9024731050
Email: research@dgiclinical.com

Abstract

Background: SymptomGuide Dementia (DGI Clinical Inc) is a publicly available online symptom tracking tool to support caregivers of persons living with dementia. The value of such data are enhanced when the specific dementia stage is identified.

Objective: We aimed to develop a supervised machine learning algorithm to classify dementia stages based on tracked symptoms.

Methods: We employed clinical data from 717 people from 3 sources: (1) a memory clinic; (2) long-term care; and (3) an open-label trial of donepezil in vascular and mixed dementia (VASPECT). Symptoms were captured with SymptomGuide Dementia. A clinician classified participants into 4 groups using either the Functional Assessment Staging Test or the Global Deterioration Scale as mild cognitive impairment, mild dementia, moderate dementia, or severe dementia. Individualized symptom profiles from the pooled data were used to train machine learning models to predict dementia severity. Models trained with 6 different machine learning algorithms were compared using nested cross-validation to identify the best performing model. Model performance was assessed using measures of balanced accuracy, precision, recall, Cohen κ, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). The best performing algorithm was used to train a model optimized for balanced accuracy.

Results: The study population was mostly female (424/717, 59.1%), older adults (mean 77.3 years, SD 10.6, range 40-100) with mild to moderate dementia (332/717, 46.3%). Age, duration of symptoms, 37 unique dementia symptoms, and 10 symptom-derived variables were used to distinguish dementia stages. A model trained with a support vector machine learning algorithm using a one-versus-rest approach showed the best performance. The correct dementia stage was identified with 83% balanced accuracy (Cohen κ=0.81, AUPRC 0.91, AUROC 0.96). The best performance was seen when classifying severe dementia (AUROC 0.99).

Conclusions: A supervised machine learning algorithm exhibited excellent performance in identifying dementia stages based on dementia symptoms reported in an online environment. This novel dementia staging algorithm can be used to describe dementia stage based on user-reported symptoms. This type of symptom recording offers real-world data that reflect important symptoms in people with dementia.

(J Med Internet Res 2020;22(11):e20840) doi:10.2196/20840

Introduction

Background

People living with dementia experience a variety of symptoms. These symptoms cross several domains beyond cognition, including executive function (eg, planning [1]), behavior (eg, agitation [2]), and physical manifestations (eg, mobility [3]). This heterogeneity of symptoms is further increased by changes in daily occurrence and manifestation. Furthermore, these combinations can vary both between people and within people across time [4-6]. This variability can be informative. Hallucinations, for example, have been reported in all stages of Alzheimer disease but most commonly at later stages [7]. In contrast, in people with Lewy body dementia, they can be a presenting feature [8]. The complex nature of dementia poses diagnostic and management challenges for health care professionals [9,10]. A key strategy is recognizing patterns, which forms the basis of dementia staging. Pattern recognition can be enhanced by tracking dementia symptoms early in the course of progressive cognitive impairment. This is especially useful when employing an approach that allows common but under-studied symptoms (eg, verbal repetition [11] or misplacing objects [12]), which may nevertheless be informative when assembled in an accessible fashion [5,6] or respond to treatment [13], to be recognized and evaluated.

Requirement for Dementia Staging Tools

To allow individual applicability, any treatment approach must consider the person's dementia stage [14]. Several clinician-facilitated dementia tools allow face-to-face staging, including the Global Deterioration Scale (GDS) [15], the Functional Assessment Staging Test (FAST) [16], the Dependence Scale [17], and the Clinical Dementia Rating Scale Sum of Boxes [18]. Defining dementia from unadjudicated online encounters (ie, where people living with dementia symptoms or their care partners track their symptoms in a web-based tool) is an important challenge that could improve both early detection and treatment evaluation [19]. Even so, dementia staging from solely online interactions has rarely been explored [20-22].

Online symptom tracking tools are common ways to help health care professionals understand dementia symptoms. They can also be valuable as education tools. SymptomGuide Dementia (DGI Clinical Inc) is an online dementia symptom tracking tool that provides a library of common and distressing symptoms. It serves as an educational tool and allows a user to identify symptoms of concern and track their change over time [5,23]. Earlier, we developed an algorithm to stage dementia severity into 4 levels of cognitive impairment for use with SymptomGuide Dementia or other similar databases using clinician-staged symptom profiles of 320 people [24]. Here, we aimed to develop a new staging algorithm using machine learning techniques with training data from a larger and more diverse set of clinical data and to validate this approach with well-established clinical dementia staging tools.

Methods

Participants and Procedure

Data for this study were obtained from a tertiary care memory clinic in Halifax, Nova Scotia, Canada from 2007 to 2013 as well as data from a study in long-term care, and an open-label trial of donepezil in vascular and mixed dementia (VASPECT) clinical trial [25,26]. Data from patients and family members (care partners) were collected using SymptomGuide Dementia in its electronic (web-based) or paper format. In addition, participants (N=717) underwent standard clinical assessments, including staging of dementia with one of two clinical tests, the GDS or the FAST. Both GDS and FAST have excellent reliability and validity [16,27]. Additionally, FAST stages have been shown to be concordant with GDS stages, and a correlation of 0.9 has been observed between them [28]. The GDS and FAST scores were interpreted as follows: a score of 3 indicated mild cognitive impairment, a score of 4 indicated mild dementia, a score of 5 indicated moderate dementia, and a score of 6 indicated severe dementia. These stages were used as target variables for classification prediction. All 4 stages were treated as discrete; therefore, discriminative models were used to perform the classification task. Only data collected at baseline (first visit) for each participant were prepared and used to train the models.

A web-based symptom tracking tool aimed to support caregivers of persons living with dementia, SymptomGuide Dementia, was used for data capture and storage for data obtained from the 3 sources. The symptoms can be either selected from an existing library of standardized symptoms or created by the caregiver. For each of the standardized symptoms, several plain-language descriptors are present. These provided another submenu for selection by the user. For each symptom selected, users were asked to indicate the frequency of the symptom and rank all the symptoms from most to least important. Users were also asked to input demographic information (eg, age and gender) and health-related information (eg, duration since first symptom), which was attached to their symptom profiles. Symptom information for each participant in the 3 sources was coded in the same format as represented in the online database. We, therefore, refer to *participants* when describing their characteristics and *user profiles* in relation to the representation of their symptoms.

Data Preparation

Users who did not select at least one symptom from the existing library of standardized symptoms were excluded from the analysis. Any patient age reported as less than 40 years was replaced with the group average for the respective stage. This was done with the assumption that the survey question was misinterpreted, and the reported age was the care partner's age not the age of the participant with dementia. Each symptom was represented by the ratio of descriptors selected for that symptom to the total number of descriptors selected across all symptoms by the participant. In addition to individual

symptoms, the ratio of selected descriptors and ratio of reported frequency of all symptoms were grouped into the following 5 domains: Behavioral Function, Cognitive Function, Daily Function, Executive Function, and Physical Manifestations for each participant [29]. Finally, age and the duration of symptoms were also included as features (variables used for prediction). All features were continuous except for the duration of symptoms which was treated as categorical data (I don't know, 1-3 months, 3-6 months, 6-12 months, 1-2 years, 1 year or more). Of the 56 symptoms in the standardized menu in SymptomGuide Dementia, 37 symptoms were selected to be included in the final set of features. This was accomplished by pruning symptoms based on a minimum occurrence of at least 15 times. As before in the algorithm developed with 320 users [24], here we maintained the 4 common clinical classifications of mild cognitive impairment, and mild dementia, moderate dementia, and severe dementia. Of the 717 users, a majority (332, 46.3%) were clinically staged as having mild dementia with the FAST or GDS, 133 (18.5%) as having mild cognitive impairment, 138 (19.2%) as having moderate dementia, and 114 (15.8%) as having severe dementia.

Since the different dementia stages were not equally represented in the data set, the minority stages (eg, mild cognitive impairment, moderate dementia, and severe dementia) were oversampled and the most represented stage (mild dementia) was undersampled in the machine learning pipeline. Oversampling was done with the borderline variant of the synthetic minority oversampling technique algorithm with a target of increasing the minority stage sizes by approximately 1.45 times their original size [30]. Undersampling was done with the neighborhood cleaning rule algorithm that focuses on data cleaning rather than data reduction. This technique has been previously shown to improve identification of minority classes in machine learning [31].

Building the Model

The models were adjudicated and iterated using measures of balanced accuracy, precision (also known as positive predictive value), recall (also known as sensitivity), Cohen κ, area under the receiver operating characteristics curve (AUROC), and area under the precision-recall curve (AUPRC). Balanced accuracy in this study was the average of individual accuracy for each stage [32]. In a balanced data set, this score would represent the accuracy. Data were stratified by stage and randomly split with 70% of the data used as a training data set (n=502) and 30% of the data used as a test data set (n=215) for validation.

The use of a single set of data to conduct both model selection and model training can lead to overfitting and selection bias [33]. To address this, we used a nested cross-validation approach as described in Figure 1. The average inner cross-validation estimates of the primary selection criterion were maximized by selecting optimal hyperparameters from a range of possible values. The inner and outer cross-validation loops were repeated 3 times to account for variance arising from choice of data set splits [34,35]. We used 5-fold cross validation for both the inner and the outer loops. We used balanced accuracy here as the primary selection criterion for the hyperparameter tuning in the inner loop. Balanced accuracy was also used for the outer loop to provide a measure of model performance. The following machine learning algorithms were used to train models: support vector machine, k-nearest neighbor, random forest, neural network, logistic regression, stochastic gradient boosting.

Figure 1. Pseudocode representation of the nested cross-validation procedure used during model selection trials.

Algorithm 1: Stratified K-Fold Nested Cross-Validation with Grid Search

---

Require: K_1, K_2, where K_1 is number of outer folds and K_2 inner folds
Require: D, dataset containing input features X and output feature y
Require: P_{sets}, set of hyper-parameters with different values
Require: M, a single estimator, model.
Split D into stratified K_1 folds
for $i = 1$ *to* K_1 splits **do**
    Set fold i as D_i^{test}, and remaining folds as D_i^{train}
    Split D_i^{train} into stratified K_2 folds
    for $j = 1$ *to* K_2 **do**
        Set fold j as D_j^{test}, and remaining folds as D_j^{train}
        foreach p *in SequentialSample(P_{sets})* **do**
            Train M on D_j^{train} with hyperparameter set p
            Compute test error E_j^{test} for M with D_j^{test}
        end
    end
    Select optimal hyperparameter set p^* from P_{sets}, where E_j^{test} is best
    Train M with D_i^{train}, using p^*
    Compute test error E_i^{test} for M with D_i^{test}
end

The best performing algorithm was then trained on the complete training data set with nested cross-validated hyperparameter tuning to model the data. To understand model performance and guard against overfitting, the model was tested against the test data set to obtain performance estimates of following metrics: weighted precision, weighted recall, balanced accuracy, Cohen κ, AUROC, and AUPRC. The final model was further assessed with a permutation test, which measured the likelihood of obtaining the observed accuracy by chance. This was done by repeating the classification (training and testing) procedure 200 times after randomly shuffling the data and permuting the labels in each iteration. The scores obtained with the permuted data were compared with the scores from the original data. We computed the probability of obtaining a score with permuted data that was better than with original data. Obtaining a small probability value rejects the null hypothesis that our model performed better than random chance and that the model had learned a real relationship between our data and dementia stages [36]. In other words, this process estimates how likely it is to obtain the observed classification performance on the test set by chance [37].

All data processing, analysis, and visualization were performed using Python (version 3.6; 64-bit) libraries (numpy, version 1.18.1; scipy, version 1.4.1; matplotlib, version 3.1.2; pandas, version 0.25.3) [38-42]. Classification algorithms were processed and analyzed using scikit-learn (version 0.22.1) and scipy [39,43]. The synthetic minority oversampling technique and neighborhood cleaning rule were implemented using imbalanced-learn (version 0.6.1) [44].

Results

Participants

This study used data from memory clinic (n=420), a long-term care study (n=169), and the VASPECT clinical trial (n=128) for a participant sample that allows 717 user profiles in people with clinical diagnosis and staging (Table 1) [25,26]. The mean participant age was 77.3 years (SD 10.6 years), and 59.1% of the participants were women. The mean FAST score was 4.0 (SD 0.9), and the mean GDS score was 4.8 (SD 1.9). The participants identified a median of 5 symptoms (range 1-27).

Table 1. Descriptive statistics of participants from clinical studies, by data source.

Characteristic	Memory clinic	Long-term care	VASPECT[a] open-label trial	Total
Sample size, n (%)	420 (58.5)	169 (23.5)	128 (18.0)	717 (100)
Age (in years), mean (SD)	74.6 (12.5)	81.0 (19.1)	75.4 (9.2)	77.3 (10.6)
FAST, mean (SD)	4.0 (0.9)	5.3 (1.1)	4.3 (0.5)	4.1 (0.9)
GDS, mean (SD)	4.8 (1.9)	5.2 (1.0)	—[b]	5.2 (1.1)
Reported symptoms, median (range)	5 (1-14)	4 (1-12)	6 (1-27)	5 (1-27)
Sex, n (%)				
Female	228 (54.3)	129 (76.3)	67 (52.3)	424 (59.1)
Male	192 (45.7)	40 (23.7)	61 (47.7)	293 (40.9)
FAST, mean (SD)	4.0 (0.9)	5.3 (1.1)	4.3 (0.5)	4.1 (0.9)
GDS, mean (SD)	4.8 (1.9)	5.2 (1.0)	—	5.2 (1.1)
Reported symptoms, median (range)	5 (1-14)	4 (1-12)	6 (1-27)	5 (1-27)
Reported symptoms by dementia stage, median (range)				
Mild cognitive impairment	3 (1-14)	2 (2-4)	—	4 (1-14)
Mild dementia	5 (1-11)	3 (1-8)	6 (1-24)	5 (1-24)
Moderate dementia	5.5 (1-11)	4 (1-7)	7 (2-27)	4.5 (1-27)
Severe dementia	5 (2-11)	5 (2-12)	10 (7-13)	5 (1-13)
Stage, n (%)				
Mild cognitive impairment	126 (30.0)	7 (4.1)	—	133 (18.5)
Mild dementia	203 (48.3)	33 (19.5)	96 (75)	332 (46.3)
Moderate dementia	58 (13.8)	50 (29.6)	30 (23.4)	138 (19.2)
Severe dementia	33 (7.8)	79 (46.7)	2 (1.5)	114 (15.8)
Age (years), mean (SD)				
Mild cognitive impairment	71.2 (14.3)	87.5 (9.9)	—	73.2 (12.2)
Mild dementia	75.3 (11.8)	81.5 (16.3)	74.7 (8.8)	76.4 (9.8)
Moderate dementia	77.1 (10.1)	83.6 (14.1)	77.5 (9.9)	80 (9.4)
Severe dementia	77.3 (10.7)	78.5 (23)	75 (16.9)	81.2 (9.7)

[a]An open-label trial of donepezil in vascular and mixed dementia.

[b]No data from this source.

Symptoms

Table 2 shows the frequency of dementia symptoms reported for user profiles and classified by dementia stage as assessed clinically with the FAST and GDS tools.

Table 2 illustrates the relationship between symptom frequency and clinical dementia stage. There was a sharp increase in the frequency of aggression, wandering, and incontinence in patients with severe dementia. By contrast, symptoms such as memory of recent events, repetitive questioning, and initiative declined with increasing dementia severity.

Table 2. Mean frequency of reported user profile symptoms of clinical study participants by clinically defined dementia stage.

Symptom	Mild cognitive impairment, n (%)	Mild dementia, n (%)	Moderate dementia, n (%)	Severe dementia, n (%)
Aggression	1 (0.8)	7 (2.1)	4 (2.9)	32 (28.1)
Anxiety & worry	35 (26.3)	82 (24.7)	23 (16.7)	19 (16.7)
Appetite	0 (0)	28 (8.4)	12 (8.7)	9 (7.9)
Balance	5 (3.8)	7 (2.1)	13 (9.4)	19 (16.7)
Bathing	0 (0)	9 (2.7)	11 (8.0)	6 (5.3)
Delusions & paranoia	4 (3.0)	29 (8.7)	20 (14.5)	19 (16.7)
Disorientation to place	3 (2.3)	28 (8.4)	15 (10.9)	18 (15.8)
Disorientation to time	4 (3.0)	41 (12.3)	25 (18.1)	23 (20.2)
Dressing	1 (0.8)	13 (3.9)	15 (10.9)	7 (6.1)
Eating	0 (0)	5 (1.5)	3 (2.2)	9 (7.9)
Financial management	5 (3.8)	30 (9.0)	11 (8.0)	1 (0.9)
Following instructions	6 (4.5)	12 (3.6)	2 (1.4)	2 (1.8)
Hallucinations	0 (0)	15 (4.5)	11 (8.0)	12 (10.5)
Hobbies & games	2 (1.5)	17 (5.1)	5 (3.6)	0 (0)
Household chores	4 (3.0)	39 (11.7)	13 (9.4)	5 (4.4)
Incontinence	1 (0.8)	6 (1.8)	6 (4.3)	23 (20.2)
Insight	4 (3.0)	19 (5.7)	6 (4.3)	2 (1.8)
Interest initiative	46 (34.6)	148 (44.6)	39 (28.3)	19 (16.7)
Irritability frustration	19 (14.3)	92 (27.7)	24 (17.4)	28 (24.6)
Judgment	13 (9.8)	42 (12.7)	24 (17.4)	25 (21.9)
Language difficulty	18 (13.5)	59 (17.8)	15 (10.9)	21 (18.4)
Low mood	3 (2.3)	44 (13.3)	13 (9.4)	6 (5.3)
Meal preparation cooking	10 (7.5)	57 (17.2)	19 (13.8)	2 (1.8)
Memory for names faces	2 (1.5)	29 (8.7)	24 (17.4)	12 (10.5)
Memory of future events	0 (0)	37 (11.1)	16 (11.6)	1 (0.9)
Memory of past events	8 (6.0)	33 (9.9)	21 (15.2)	13 (11.4)
Memory of recent events	100 (75.2)	233 (70.2)	81 (58.7)	30 (26.3)
Misplacing or losing objects	21 (15.8)	53 (16.0)	11 (8)	5 (4.4)
Mobility	4 (3.0)	14 (4.2)	16 (11.6)	30 (26.3)
Operating gadgets/appliances	8 (6.0)	76 (22.9)	24 (17.4)	4 (3.5)
Personal care hygiene	7 (5.3)	26 (7.8)	36 (26.1)	28 (24.6)
Physical complaints	2 (1.5)	15 (4.5)	7 (5.1)	4 (3.5)
Repetitive questions stories	51 (38.3)	169 (50.9)	50 (36.2)	18 (15.8)
Shadowing	1 (0.8)	8 (2.4)	8 (5.8)	4 (3.5)
Social interaction/withdrawal	20 (15.0)	64 (19.3)	22 (15.9)	11 (9.6)
Wandering	0 (0)	2 (0.6)	10 (7.2)	29 (25.4)

Model Selection

Six machine learning models were tested on the training data set. Table 3 illustrates the models used and the validation data obtained for each model in terms of accuracy, precision, and recall when predicting dementia stage. The table also indicates

values for the Cohen κ, which measures the agreement between the dementia stage predicted by the model and the dementia stage as determined clinically. At the end of the model selection process, the model trained with a support vector machine was selected as the best performing model when used with the training data set (Table 3).

Table 3. Performance of candidate models with the training data set.

Model	Balanced accuracy, mean (SD)	Precision (weighted), mean (SD)	Recall (weighted), mean (SD)	Cohen κ, mean (SD)
Support vector machine	0.73 (0.07)	0.75 (0.07)	0.75 (0.06)	0.65 (0.09)
k-nearest neighbor	0.72 (0.08)	0.73 (0.07)	0.72 (0.07)	0.62 (0.10)
Random forest	0.70 (0.07)	0.74 (0.08)	0.73 (0.07)	0.62 (0.09)
Neural network	0.66 (0.10)	0.67 (0.11)	0.66 (0.09)	0.54 (0.13)
Logistic regression	0.65 (0.08)	0.66 (0.08)	0.66 (0.08)	0.53 (0.10)
Gradient boosting	0.68 (0.07)	0.70 (0.07)	0.70 (0.07)	0.58 (0.10)

Next, the support vector machine was trained and optimized with a nested cross-validated grid search on the complete training set. The final trained model was used with the test data set to obtain performance metrics for this new data subset (balanced accuracy 0.85; AUROC 0.96, weighted precision 0.87; weighted recall 0.86; AUPRC 0.91), indicating excellent model performance.

Final Model Prediction Based on Dementia Stage

The ability of the support vector machine model to predict each of the 4 dementia stages showed excellent precision and recall for all dementia stages (Table 4).

To better demonstrate predictions made across the dementia stages by the model, a confusion matrix is presented in Figure 2.

To determine the relationship between the true positives and false positives identified by the model, receiver operating characteristic curves of the model's output were plotted (Figure 3). The AUROC for the overall model was high (AUROC 0.96). The final model achieved the best results when classifying severe dementia (AUROC 0.98) and mild cognitive impairment (AUROC 0.97).

Table 4. Precision and recall of model prediction by dementia stage.

Stage	Precision	Recall
Mild cognitive impairment	0.85	0.87
Mild dementia	0.82	0.89
Moderate dementia	0.91	0.80
Severe dementia	0.93	0.86

XSL•FO
RenderX

Figure 2. Confusion matrix of the trained model. Each row of the matrix represents the instances of actual dementia stage while each column represents the instances of predicted dementia stage. Counts are colored from the highest cell (darker) to the lowest (lighter). The top-left to bottom-right diagonal cells count correctly predicted dementia stages.
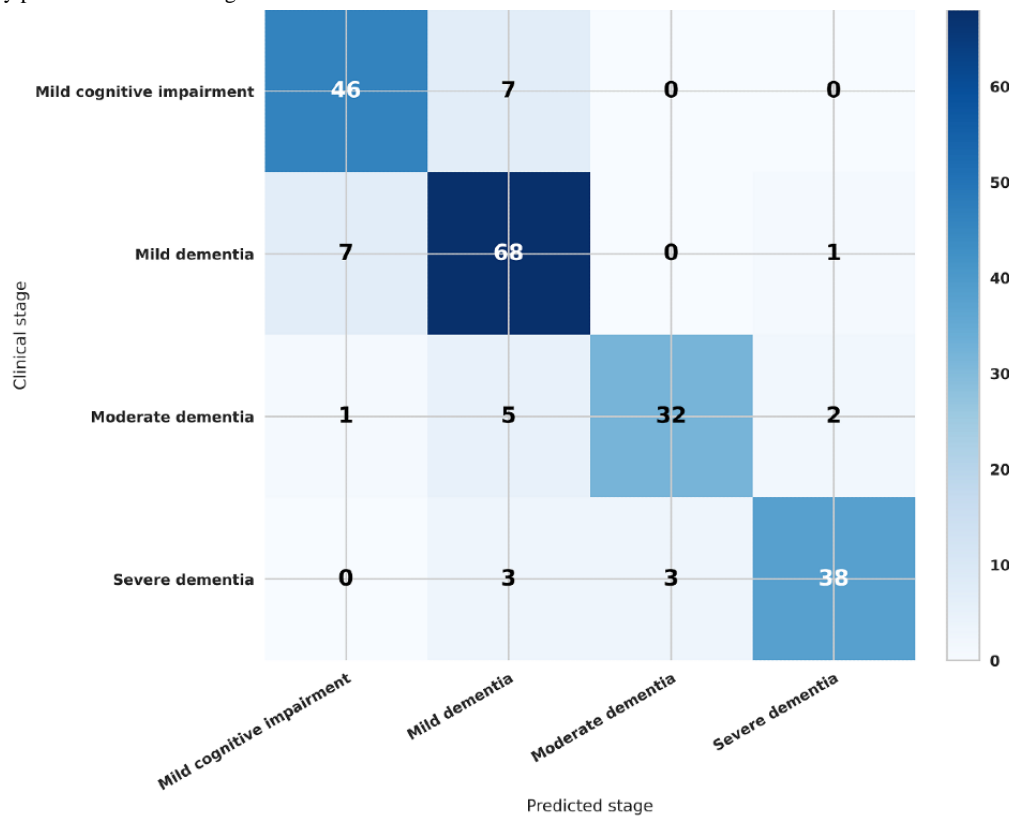
Figure 3. Receiver operating characteristic curves for each dementia stage predicted by the model. AUC: area under the curve; ROC: receiver operating characteristics.
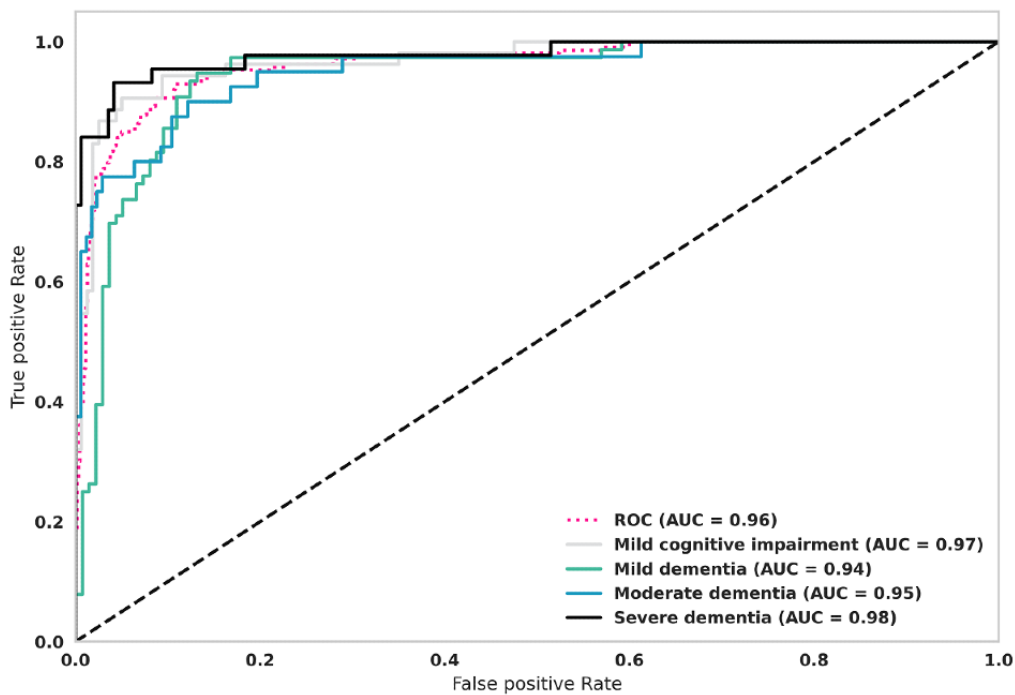
Another way to assess the relationship between false positives and false negatives is to use a precision-recall curve, where high precision indicates a low false positive rate, and high recall denotes a low false negative rate. Figure 4 shows precision-recall curves of the overall model output by dementia stage. The overall model performed well (AUPRC 0.91). When AUPRC metrics were compared for individual dementia stages, the model performed best when classifying severe dementia (AUPRC 0.95) and mild cognitive impairment (AUPRC 0.93). It was somewhat less able to discriminate between mild and moderate

dementia (Figure 4). These observations are similar to those seen when these relationships were evaluated with receiver operating characteristic curves as shown in Figure 3.

To confirm that the model could accurately predict dementia stage, we performed a permutation test, where we used randomly mislabeled data in several iterations, grouped about the level expected by chance (Figure 5). The random permutation scores had a balanced accuracy between 0.2 and 0.3. This was well short of the classification score for the actual data, which had a balanced accuracy of 0.85, and a probability of obtaining this by chance <.005.

Figure 4. Precision-recall curves for each dementia stage as predicted by the model. AUC: area under the curve; PRC: precision-recall curve.
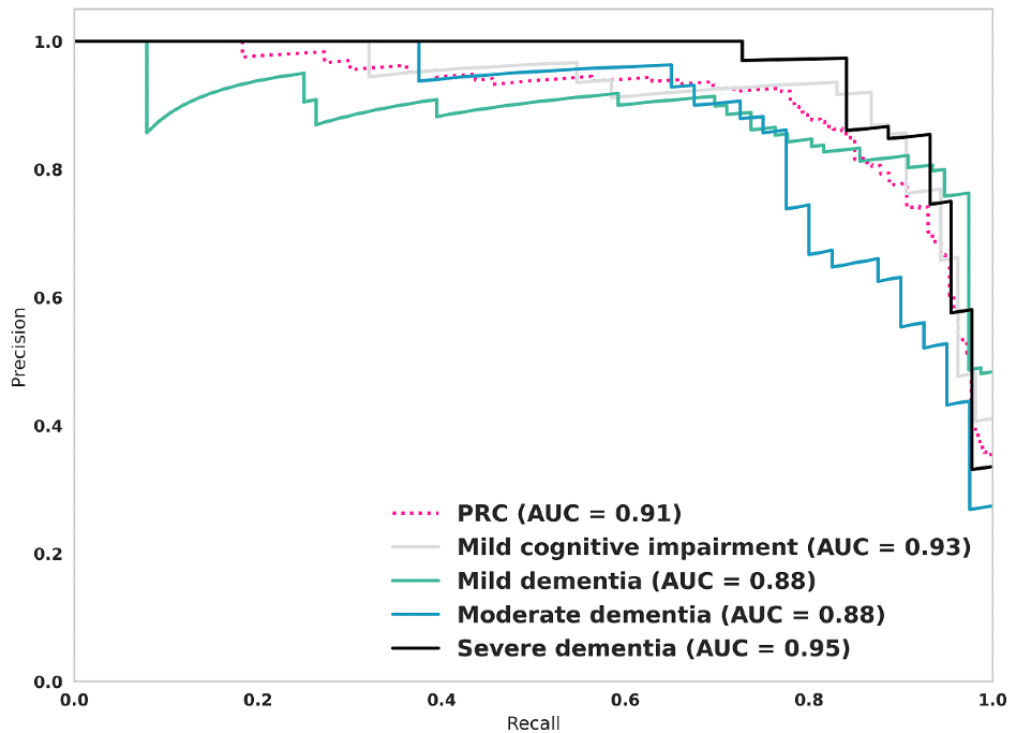
Figure 5. The classification scores obtained from models trained on permuted data were well short of scores obtained with the model trained on original data.
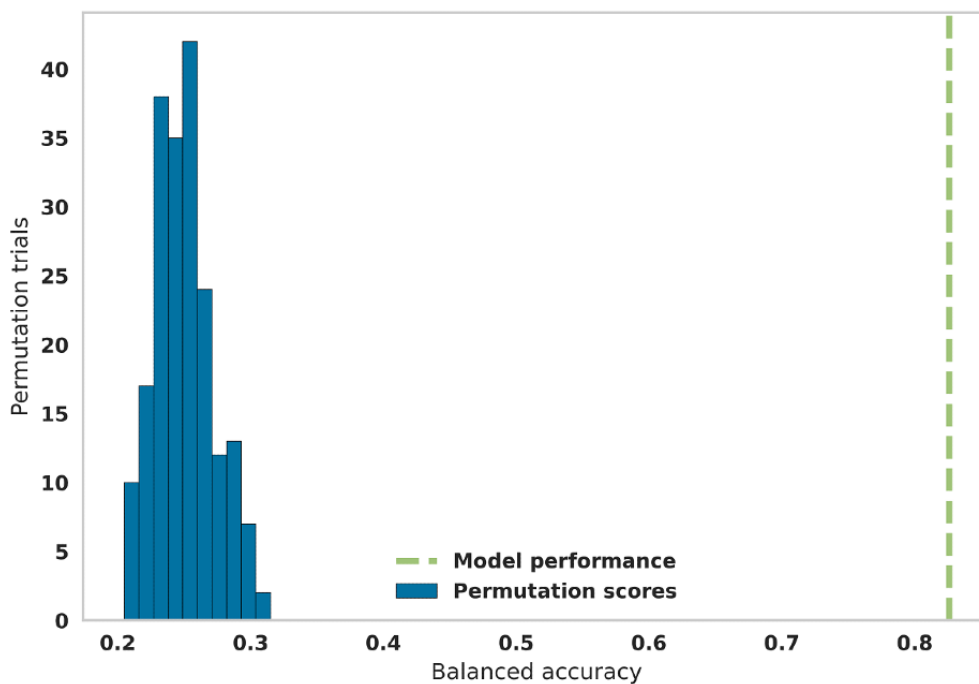
Discussion

This study aimed to stage dementia severity based on symptom profiles constructed with a standardized symptom menu from an online symptom tracking tool. We found that a support vector machine model consistently predicted each of the 4 dementia stages based on online symptom data reported by caregivers of persons with dementia. This approach to staging dementia severity will allow us to gain insights from online reported symptom data that can be collected by SymptomGuide Dementia and other similar platforms. In this way, symptom reporting can facilitate understanding dementia progression. For example, earlier work from this database suggests important qualitative differences in symptoms such as misplacing objects (eg, with dementia progression, less instances of simply forgetting where an item might be and more instances of placing items in an odd place [6]) or verbal repetition (eg, repetitive questioning, most often seen in mild dementia; it is characteristically dementia-defining when seen with early functional decline—difficulty operating familiar gadgets or appliances [5]). In this way, allowing the patients and carer voices to contribute to our understanding of dementia phenomenology can lead to recognizing patterns of both progression—as above—and of treatment [13]. The updated staging algorithm described here will further such inquiries.

We trained multiple machine learning algorithms and selected the best performing algorithm to use for our dementia stage classification task. A support vector machine model using a one-versus-rest approach demonstrated the best performance during model selection trials. The selected algorithm was then trained on the complete training data and validated using a test data set. The final model demonstrated excellent performance in discriminating dementia stages (balanced accuracy 0.85, AUROC 0.96). Receiver operating characteristic curves tend to present an optimistic picture of performance when the data set has a skewed distribution of the target variable [45]. For this reason, the performance of the model was also assessed with precision-recall curves. These too demonstrated that the model performed well, especially when classifying severe dementia. Since mild cognitive impairment and severe dementia can be considered bookends to the dementia spectrum, we can be reassured of both the model's precision and recall in classifying these extremes. For example, our model correctly classified a 75-year-old participant who reported 4 symptoms (social interaction/withdrawal, irritability and frustration, interest and initiative, aggression) as having mild cognitive impairment and a 76-year-old participant as having severe dementia based on a different set of 3 symptoms (wandering, delusions and paranoia, and aggression). The model was somewhat less accurate when classifying mild and moderate dementia. This is perhaps not surprising as symptom profiles in the middle of the dementia spectrum can exhibit a higher degree of overlap and can be difficult to distinguish clinically as well [46].

The very low probability value from the permutation tests (<.005) reassures us that the model learned a real relationship between the data and dementia stages. It demonstrates that the classification performance of the model with respect to the test set is unlikely to have occurred as a result of chance.

Our data must be interpreted with caution. For model stability, symptoms were eliminated based on a set threshold of occurrence. While this worked well here, it might not hold in a larger data set. In addition, we used 3 separate data sets that used variations of our standardized symptom menu, with differences in the composition and order of presentation of the symptoms. Since most of these patient symptom profiles were constructed with the supervision of a clinician or a rater, the model may be less generalizable to web-based symptom profiles constructed without clinician facilitation or guidance.

Several other recent studies have applied machine learning algorithms for dementia research [47-51]. Most have used neuroimaging or biomarker data to train these models. Most models trained with neuroimaging data focus on distinguishing individual patients from healthy controls, whereas our model distinguished between different stages of dementia severity [52]. Extraction of image characteristics from neuroimaging data can be susceptible to variations in the scanner hardware and image acquisition protocols. This can produce models that may not be generalizable when applied to data acquired from different imaging sources [52]. Additionally, scans such as amyloid positron emission tomography imaging, used for diagnostic certainty regarding Alzheimer disease, can cost upward of US $4000. Machine learning models that do not rely neuroimaging data to stage or diagnose dementia, if used clinically, can potentially reduce the number of participants that require expensive neuroimaging tests [53].

More recent studies have also used data extracted from electronic health records which may include structured and unstructured data such as clinical notes, drug prescriptions, and diagnosis codes to develop predictive models [54-60]. These models have been trained to predict future onset of dementia [53-56] or diagnose undetected dementia [57,58,60] with varying levels of accuracy and can potentially serve as case-finding algorithms to target high-risk patients with further clinical assessments to confirm dementia diagnosis [58]. However, these models are contingent on the availability of consolidated electronic health records, sufficient health care interactions by the patient, and correctly transcribed notes and diagnosis codes [55,57,61]. In contrast, the model developed here does not use data extracted from electronic health records, rather it predicts dementia severity based on self-reported caregiver data and can be used to potentially unlock insights from online self-reported symptoms.

Few studies have used machine learning models to stage the severity of dementia or differentiate types of dementia [62]. One such study uses a combination of cognitive function tests and clinicians' assessments of patients to assess dementia severity on the Clinical Dementia Rating Scale [63]. On the other hand, a combination of neuropsychiatric assessment, mental status examination, and laboratory investigations have also been used to classify dementia severity with a high degree of accuracy [64]. Such approaches require trained interviewers and clinician assessment to obtain input data for the predictive models. This is in contrast to the model developed here, which is designed to stage dementia severity based on self-reported data thereby potentially offering a more economically viable screening tool for dementia severity.

Even though our sample size (n=717) is relatively small, it is larger than that of other machine learning studies in dementia, except for a 2019 report that used administrative data to diagnose incident dementia [47]. The advantage of utilizing patient reported outcomes such as SymptomGuide is that it reflects the lived experience of the patient or caregiver and focuses on what is meaningful to them. It is easier to source and computationally less expensive to train models when compared to imaging data or complex biomarkers [48,49]. Interestingly, Chiu et al [50] reported that a machine learning algorithm could be used to derive a screening instrument to distinguish normal cognition, mild cognitive impairment, and dementia. This further emphasizes that dementia symptoms can be used with machine learning to characterize various stages of dementia. On the other hand, our approach used a patient-derived library of symptoms to train a machine learning model, whereas Chui et al used machine learning to reduce the dimensionality of their screening instrument [50]. It is likely that, given the high dimensionality of late-life dementia, different machine learning approaches may be useful in dementia research. In our earlier work, we developed a model based on a neural network trained on 320 symptom profiles reported by caregivers of persons with dementia [24]. This study expands on our previous work by increasing the sample size and diversity of the training data. We also examined the performance of multiple machine learning algorithms on the available data to maximize our interpretation. The support vector machine outperformed the neural network approach, highlighting the advantage of the current approach.

Future studies could include integrating the model developed here with an electronic interface by which end users could build a symptom profile and obtain the dementia stage. This instrument also has the potential to facilitate physician-patient discussions or to aid screening patients before their in-person memory clinic visit. This model can potentially be applied on other web-based data sets that contain symptom profiles of persons affected with dementia.

The model presented here can classify dementia stages from individualized symptom data. This real-world evidence will enable us to better understand the symptoms that matter most to people affected by dementia at each dementia stage. That information can greatly expand access to understanding the lived experience of dementia.

Conflicts of Interest

References

1. Kirova A, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. Biomed Res Int 2015;2015:748212 [FREE Full text] [doi: 10.1155/2015/748212] [Medline: 26550575]

2. Rockwood K, Sanon Aigbogun M, Stanley J, Wong H, Dunn T, Chapman CAT, et al. The symptoms targeted for monitoring in a web-based tracking tool by caregivers of people with dementia and agitation: cross-sectional study. J Med Internet Res 2019 Jun 28;21(6):e13360 [FREE Full text] [doi: 10.2196/13360] [Medline: 31254339]

3. Auer SR, Höfler M, Linsmayer E, Beránková A, Prieschl D, Ratajczak P, et al. Cross-sectional study of prevalence of dementia, behavioural symptoms, mobility, pain and other health parameters in nursing homes in Austria and the Czech Republic: results from the DEMDATA project. BMC Geriatr 2018 Aug 13;18(1):178 [FREE Full text] [doi: 10.1186/s12877-018-0870-8] [Medline: 30103672]

4. Rockwood K, Mitnitski A, Richard M, Kurth M, Kesslak P, Abushakra S. Neuropsychiatric symptom clusters targeted for treatment at earlier versus later stages of dementia. Int J Geriatr Psychiatry 2015 Apr;30(4):357-367. [doi: 10.1002/gps.4136] [Medline: 24798635]

5. Reeve E, Molin P, Hui A, Rockwood K. Exploration of verbal repetition in people with dementia using an online symptom-tracking tool. Int Psychogeriatr 2017 Dec;29(6):959-966 [FREE Full text] [doi: 10.1017/S1041610216002180] [Medline: 28274302]

6. McGarrigle L, Howlett SE, Wong H, Stanley J, Rockwood K. Characterizing the symptom of misplacing objects in people with dementia: findings from an online tracking tool. Int Psychogeriatr 2019 Nov;31(11):1635-1641. [doi: 10.1017/S104161021800220X] [Medline: 30698122]

7. El Haj M, Roche J, Jardri R, Kapogiannis D, Gallouj K, Antoine P. Clinical and neurocognitive aspects of hallucinations in Alzheimer's disease. Neurosci Biobehav Rev 2017 Dec;83:713-720 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.02.021] [Medline: 28235545]

8. Foguem C, Manckoundia P. Lewy body disease: clinical and pathological. Curr Neurol Neurosci Rep 2018 Apr 08;18(5):24. [doi: 10.1007/s11910-018-0835-5] [Medline: 29629495]

9. Woods B, Arosio F, Diaz A, Gove D, Holmerová I, Kinnaird L, et al. Timely diagnosis of dementia? family carers' experiences in 5 European countries. Int J Geriatr Psychiatry 2019 Jan;34(1):114-121 [FREE Full text] [doi: 10.1002/gps.4997] [Medline: 30246266]

10.   Knopman DS, Petersen RC. Mild cognitive impairment and mild dementia: a clinical perspective. Mayo Clin Proc 2014 Oct;89(10):1452-1459 [FREE Full text] [doi: 10.1016/j.mayocp.2014.06.019] [Medline: 25282431]

11.   Cook C, Fay S, Rockwood K. Verbal repetition in people with mild-to-moderate Alzheimer Disease: a descriptive analysis from the VISTA clinical trial. Alzheimer Dis Assoc Disord 2009;23(2):146-151. [doi: 10.1097/WAD.0b013e318193cbef] [Medline: 19474572]

12.   Hamilton L, Fay S, Rockwood K. Misplacing objects in mild to moderate Alzheimer's disease: a descriptive analysis from the VISTA clinical trial. J Neurol Neurosurg Psychiatry 2009 Sep;80(9):960-965. [doi: 10.1136/jnnp.2008.166801] [Medline: 19293172]

13.   Rockwood K, Fay S, Jarrett P, Asp E. Effect of galantamine on verbal repetition in AD: a secondary analysis of the VISTA trial. Neurology 2007 Apr 3;68(14):1116-1121. [doi: 10.1212/01.wnl.0000258661.61577.b7] [Medline: 17404193]

14.   Lopez OL, Becker JT, Sweet RA, Klunk W, Kaufer DI, Saxton J, et al. Psychiatric symptoms vary with the severity of dementia in probable Alzheimer's disease. JNP 2003 Aug;15(3):346-353. [doi: 10.1176/jnp.15.3.346]

15.   Reisberg B, Ferris S, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982 Sep;139(9):1136-1139. [doi: 10.1176/ajp.139.9.1136] [Medline: 7114305]

16.   Sclan SG, Reisberg B. Functional Assessment Staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. Int. Psychogeriatr 2005 Jan 07;4(3):55-69. [doi: 10.1017/s1041610292001157]

17.   Stern Y, Albert SM, Sano M, Richards M, Miller L, Folstein M, et al. Assessing patient dependence in Alzheimer's disease. Journal of Gerontology 1994 Sep 01;49(5):M216-M222. [doi: 10.1093/geronj/49.5.m216]

18.   Hughes CP, Berg L, Danziger W, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 2018 Jan 29;140(6):566-572. [doi: 10.1192/bjp.140.6.566]

19.   Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits With Google Trends: evidence from Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264. [doi: 10.2196/jmir.4516]

20.   Snyder PJ, Kahle-Wrobleski K, Brannan S, Miller DS, Schindler RJ, DeSanti S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? Alzheimer's & Dementia 2014 Nov 01;10(6):853-860. [doi: 10.1016/j.jalz.2014.07.158]

21.   Yuan Q, Tan TH, Wang P, Devi F, Ong HL, Abdin E, et al. Staging dementia based on caregiver reported patient symptoms: Implications from a latent class analysis. PLoS ONE 2020 Jan 15;15(1):e0227857. [doi: 10.1371/journal.pone.0227857]

22.   Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]

23.   Rockwood K. An individualized approach to tracking and treating Alzheimer's disease. Clin Pharmacol Ther 2010 Oct;88(4):446-449. [doi: 10.1038/clpt.2010.68]

24.   Rockwood K, Richard M, Leibman C, Mucha L, Mitnitski A. Staging dementia from symptom profiles on a care partner website. J Med Internet Res 2013 Aug 07;15(8):e145. [doi: 10.2196/jmir.2461]

25.   Rockwood JK, Richard M, Garden K, Hominick K, Mitnitski A, Rockwood K. Precipitating and predisposing events and symptoms for admission to assisted living or nursing home care. Can Geriatr J 2014 Mar;17(1):16-21 [FREE Full text] [doi: 10.5770/cgj.17.93] [Medline: 24596590]

26.   Rockwood K, Mitnitski A, Black SE, Richard M, Defoy I, VASPECT study investigators. Cognitive change in donepezil treated patients with vascular or mixed dementia. Can J Neurol Sci 2013 Jul;40(4):564-571. [doi: 10.1017/s0317167100014670] [Medline: 23786741]

27.   Foster JR, Sclan S, Welkowitz J, Boksay I, Seeland I. Psychiatric assessment in medical long-term care facilities: Reliability of commonly used rating scales. Int J Geriat Psychiatry 1988 Jul;3(3):229-233. [doi: 10.1002/gps.930030310]

28.   Auer S, Reisberg B. The GDS/FAST staging system. Int Psychogeriatr 1997;9 Suppl 1:167-171. [doi: 10.1017/s1041610297004869] [Medline: 9447440]

29.   Richard M, Rockwood K, Mitnitski A. O2-13-06: Symptom profiles in relation to dementia staging. Alzheimer's & Dementia 2012 Jul 01;8(4S_Part_7):P263-P263. [doi: 10.1016/j.jalz.2012.05.701]

30.   Han H, Wang WY, Mao BH. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: Proceedings of the 2005 International Conference on Advances in Intelligent Computing - Volume Part I. Berlin, Heidelberg: Springer-Verlag; 2005 Presented at: ICIC'05; August 23-26, 2005; Hefei, China p. 878-887. [doi: 10.1007/11538059_91]

31.   Laurikkala J. Improving Identification of Difficult Small Classes by Balancing Class Distribution. In: Proceedings of the 8th Conference on AI in Medicine in Europe: Artificial Intelligence Medicine.: Springer; 2001 Jun Presented at: Artificial Intelligence in Medicine in Europe. AIME; July 1-4, 2001; Cascais, Portugal p. 63-66. [doi: 10.1007/3-540-48229-6_9]

32.   Mosley L. A balanced approach to the multi-class imbalance problem. Graduate Theses and Dissertations Iowa State University. Ames, Iowa; 2013. URL: https://lib.dr.iastate.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=4544&context=etd [accessed 2020-11-06]

33.   Cawley GC, Talbot NL. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research 2010 Mar 01;11:2079-2107. [doi: 10.5555/1756006.1859921]

34.   Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014 Mar 29;6(1):10 [FREE Full text] [doi: 10.1186/1758-2946-6-10] [Medline: 24678909]

35. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006;7:91 [FREE Full text] [doi: 10.1186/1471-2105-7-91] [Medline: 16504092]

36. Ojala M, Garriga G. Permutation tests for studying classifier performance. In: Journal of Machine Learning Research. USA: IEEE Computer Society; 2009 Dec Presented at: ICDM IEEE International Conference on Data Mining; Dec. 6 - Dec. 9 2009; Miami, Florida p. 908-913. [doi: 10.1109/ICDM.2009.108]

37. Golland P, Fischl B. Permutation tests for classification: towards statistical significance in image-based studies. Inf Process Med Imaging 2003 Jul;18:330-341. [doi: 10.1007/978-3-540-45087-0_28] [Medline: 15344469]

38. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng 2011 Mar;13(2):22-30. [doi: 10.1109/MCSE.2011.37]

39. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Feb 3;17(3):261-272. [doi: 10.1038/s41592-019-0686-2]

40. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007 May;9(3):90-95. [doi: 10.1109/mcse.2007.55]

41. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: Python in Science Conference; June 28 - July 3; Austin, Texas p. 51-56. [doi: 10.25080/majora-92bf1922-00a]

42. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.

43. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp and Comm 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]

44. Lemaître G, Nogueira F, Aridas CK. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J Mach. Learn. Res 2017 Jan;18(1):559-563 [FREE Full text]

45. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]

46. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet 2006 Apr 15;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]

47. Nori VS, Hane CA, Crown WH, Au R, Burke WJ, Sanghavi DM, et al. Machine learning models to predict onset of dementia: A label learning approach. Alzheimers Dement (N Y) 2019;5:918-925 [FREE Full text] [doi: 10.1016/j.trci.2019.10.006] [Medline: 31879701]

48. Zheng Y, Guo H, Zhang L, Wu J, Li Q, Lv F. Machine learning-based framework for differential diagnosis between vascular dementia and Alzheimer's disease using structural MRI features. Front Neurol 2019;10:1097 [FREE Full text] [doi: 10.3389/fneur.2019.01097] [Medline: 31708854]

49. Shigemizu D, Akiyama S, Asanomi Y, Boroevich KA, Sharma A, Tsunoda T, et al. A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data. BMC Med Genomics 2019 Oct 30;12(1):150 [FREE Full text] [doi: 10.1186/s12920-019-0607-3] [Medline: 31666070]

50. Chiu P, Tang H, Wei C, Zhang C, Hung G, Zhou W. NMD-12: A new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. PLoS One 2019;14(3):e0213430 [FREE Full text] [doi: 10.1371/journal.pone.0213430] [Medline: 30849106]

51. Zhou T, Thung KH, Zhu X, Shen D. Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis. Mach Learn Med Imaging 2017 Sep;10541:132-140 [FREE Full text] [doi: 10.1007/978-3-319-67389-9_16] [Medline: 29376149]

52. Davatzikos C. Machine learning in neuroimaging: progress and challenges. NeuroImage 2019 Aug;197:652-656. [doi: 10.1016/j.neuroimage.2018.10.003]

53. Hane CA, Nori VS, Crown WH, Sanghavi DM, Bleicher P. Predicting onset of dementia using clinical notes and machine learning: case-control study. JMIR Med Inform 2020 Jun 03;8(6):e17819 [FREE Full text] [doi: 10.2196/17819] [Medline: 32490841]

54. Nori VS, Hane CA, Martin DC, Kravetz AD, Sanghavi DM. Identifying incident dementia by applying machine learning to a very large administrative claims dataset. PLoS ONE 2019 Jul 5;14(7):e0203246. [doi: 10.1371/journal.pone.0203246]

55. Ben Miled Z, Haas K, Black CM, Khandker RK, Chandrasekaran V, Lipton R, et al. Predicting dementia with routine care EMR data. Artif Intell Med 2020 Jan;102:101771. [doi: 10.1016/j.artmed.2019.101771] [Medline: 31980108]

56. Park JH, Cho HE, Kim JH, Wall MM, Stern Y, Lim H, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. NPJ Digit Med 2020;3:46 [FREE Full text] [doi: 10.1038/s41746-020-0256-0] [Medline: 32258428]

57. Barnes DE, Zhou J, Walker RL, Larson EB, Lee SJ, Boscardin WJ, et al. Development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. J Am Geriatr Soc 2020 Jan;68(1):103-111. [doi: 10.1111/jgs.16182] [Medline: 31612463]

58. Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. BMC Med Inform Decis Mak 2019 Jul 09;19(1):128 [FREE Full text] [doi: 10.1186/s12911-019-0846-4] [Medline: 31288818]

XSL•FO
RenderX

59. Mar J, Gorostiza A, Ibarrondo O, Cernuda C, Arrospide A, Iruin Á, et al. Validation of random forest machine learning models to predict dementia-related neuropsychiatric symptoms in real-world data. J Alzheimers Dis 2020;77(2):855-864. [doi: 10.3233/JAD-200345] [Medline: 32741825]

60. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. BMC Med Inform Decis Mak 2019 Dec 02;19(1):248 [FREE Full text] [doi: 10.1186/s12911-019-0991-9] [Medline: 31791325]

61. Zhu CW, Ornstein KA, Cosentino S, Gu Y, Andrews H, Stern Y. Misidentification of dementia in Medicare claims and related costs. J Am Geriatr Soc 2019 Feb;67(2):269-276 [FREE Full text] [doi: 10.1111/jgs.15638] [Medline: 30315744]

62. Pellegrini E, Ballerini L, Hernandez MDCV, Chappell FM, González-Castro V, Anblagan D, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. Alzheimers Dement (Amst) 2018;10:519-535 [FREE Full text] [doi: 10.1016/j.dadm.2018.07.004] [Medline: 30364671]

63. Shankle WR, Mania S, Dick MB, Pazzani MJ. Simple models for estimating dementia severity using machine learning. Stud Health Technol Inform 1998;52 Pt 1:472-476. [Medline: 10384501]

64. Joshi S, GG VS, P DS, KR V, LM P. Classification and treatment of different stages of Alzheimer's disease using various machine learning methods. Int J of Bioi Res 2010 Jun 30;2(1):44-52. [doi: 10.9735/0975-3087.2.1.44-52]

Abbreviations

AUROC: area under the receiver operating characteristic curve
AUPRC: area under the precision-recall curve
FAST: Functional Assessment Staging Test
GDS: Global Deterioration Scale

XSL•FO
RenderX

XSL•FO
RenderX